

Supplementary Information for “Cost Function Dependent Barren Plateaus in Shallow Parametrized Quantum Circuits”

M. Cerezo,^{1,2} Akira Sone,^{1,2} Tyler Volkoff,¹ Lukasz Cincio,¹ and Patrick J. Coles¹

¹*Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, USA.*

²*Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM, USA*

In this Supplementary Information, we present detailed proofs of the propositions, theorems, and corollaries presented in the manuscript “Cost Function Dependent Barren Plateaus in Shallow Parametrized Quantum Circuits”. In Supplementary Note 1 we first introduce several lemmas which will be useful in the derivation of our main results. Then, in Supplementary Note 2, and Supplementary Note 3 we respectively provide proofs for Propositions 1 and 2 of the main text. We then derive the general equations for the variance of the cost function partial derivative in Supplementary Note 4. In Supplementary Note 5 we explicitly evaluate the variance of the cost function derivative for the special case when $V(\theta)$ is given by a single layer of the Alternating Layered Ansatz.

In Supplementary Note 6, and Supplementary Note 7, we provide our proofs to Theorem 2 and Theorem 1, respectively. Where we remark that the proof of Theorem 2 comes before that of Theorem 1 since the latter builds on the former. Then, in Supplementary Note 8, we prove Corollaries 1, and 2. In Supplementary Note 9, we demonstrate that the local cost function for the quantum autoencoder is faithful.

Supplementary Note 1: Preliminaries

In this section, we present properties that allow for analytic calculation of integrals of polynomial functions over the unitary group with respect to the unique normalized Haar measure. For more details on this topic, we refer the reader to Ref. [1, 2]. In addition, to make the Supplementary Information more self-contained, we reiterate here the definition of a t -design. Consider a finite set $\{W_y\}_{y \in Y}$ (of size $|Y|$) of unitaries W_y on a d -dimensional Hilbert space, and let $P_{(t,t)}(W)$ be an arbitrary polynomial of degree at most t in the matrix elements of W and at most t in those of $W^\dagger$. Then, we say that this finite set is a t -design if [3]

$$\frac{1}{|Y|} \cdot \sum_{y \in Y} P_{(t,t)}(W_y) = \int_{U(d)} d\mu(W) P_{(t,t)}(W), \quad (1)$$

where in the right-hand side $U(d)$ denotes the unitary group of degree d . Equation (1) implies that averaging $P_{(t,t)}(W)$ over the t -design is indistinguishable from integrating over $U(d)$ with respect to the Haar distribution.

Given $W \in U(d)$ the following expressions are valid for the first two moments [1, 2]

$$\int_{U(d)} d\mu(W) w_{i,j} w_{p,k}^* = \frac{\delta_{i,p} \delta_{j,k}}{d}, \quad (2)$$

$$\begin{aligned} \int_{U(d)} d\mu(W) w_{i_1,j_1} w_{i_2,j_2} w_{i'_1,j'_1}^* w_{i'_2,j'_2}^* &= \frac{1}{d^2 - 1} (\delta_{i_1,i'_1} \delta_{i_2,i'_2} \delta_{j_1,j'_1} \delta_{j_2,j'_2} + \delta_{i_1,i'_2} \delta_{i_2,i'_1} \delta_{j_1,j'_2} \delta_{j_2,j'_1}) \\ &\quad - \frac{1}{d(d^2 - 1)} (\delta_{i_1,i'_1} \delta_{i_2,i'_2} \delta_{j_1,j'_2} \delta_{j_2,j'_1} + \delta_{i_1,i'_2} \delta_{i_2,i'_1} \delta_{j_1,j'_1} \delta_{j_2,j'_2}). \end{aligned} \quad (3)$$

All throughout this section the integration domain will be implied to be $U(d)$, and unless otherwise specified we consider W to be an operator acting on a Hilbert space $\mathcal{H}_w$ of dimension d . When $d = 2^m$, as occurs for a Hilbert space of m qubit, we adopt the symbol $\mathbf{i} = (i_1, \dots, i_m)$ to denote a bitstring of length m such that $i_1, i_2, \dots, i_m \in \{0, 1\}$. Moreover, given two bitstrings $\mathbf{i}$ and $\mathbf{j}$ we define their concatenation as $\mathbf{i} \cdot \mathbf{j} = (i_1, \dots, i_m, j_1, \dots, j_n)$.

Operators in the computational basis of m qubits can be written as

$$W = \sum_{\mathbf{i}, \mathbf{j}} w_{\mathbf{i}, \mathbf{j}} |\mathbf{i}\rangle \langle \mathbf{j}|, \quad W^\dagger = \sum_{\mathbf{i}', \mathbf{j}'} w_{\mathbf{i}', \mathbf{j}'}^* |\mathbf{j}'\rangle \langle \mathbf{i}'|, \quad A = \sum_{\mathbf{k}, \mathbf{l}} a_{\mathbf{k}, \mathbf{l}} |\mathbf{k}\rangle \langle \mathbf{l}|, \quad B = \sum_{\mathbf{q}, \mathbf{p}} b_{\mathbf{q}, \mathbf{p}} |\mathbf{q}\rangle \langle \mathbf{p}|.$$

From the previous expressions we can derive the following lemmas.

Lemma 1. Let $\{W_y\}_{y \in Y} \subset U(d)$ form a unitary t -design with $t \geq 1$, and let $A, B : \mathcal{H}_w \rightarrow \mathcal{H}_w$ be arbitrary linear operators. Then

$$\frac{1}{|Y|} \sum_{y \in Y} \text{Tr} [W_y A W_y^\dagger B] = \int d\mu(W) \text{Tr} [W A W^\dagger B] = \frac{\text{Tr} [A] \text{Tr} [B]}{d}. \quad (4)$$

Proof. The first equality follows from the definition of a t -design. Note that $\text{Tr}[WAW^\dagger B]$ can be written as

$$\text{Tr}[WAW^\dagger B] = \sum_{\mathbf{i}_1, \mathbf{j}_1, \mathbf{i}'_1, \mathbf{j}'_1} a_{\mathbf{j}_1, \mathbf{j}'_1} b_{\mathbf{i}'_1, \mathbf{i}_1} w_{\mathbf{i}_1, \mathbf{j}_1} w_{\mathbf{i}'_1, \mathbf{j}'_1}^*.$$

Then, from (2), we have

$$\int d\mu(W) \text{Tr}[WAW^\dagger B] = \frac{1}{d} \sum_{\mathbf{i}_1, \mathbf{j}_1} a_{\mathbf{j}_1, \mathbf{j}_1} b_{\mathbf{i}_1, \mathbf{i}_1} = \frac{\text{Tr}[A] \text{Tr}[B]}{d}.$$

□

Lemma 2. Let $\{W_y\}_{y \in Y} \subset U(d)$ form a unitary t -design with $t \geq 2$ and let $A, B, C, D : \mathcal{H}_w \rightarrow \mathcal{H}_w$ be arbitrary linear operators. Then

$$\begin{aligned} \frac{1}{|Y|} \sum_{y \in Y} \text{Tr}[W_y A W_y^\dagger B W_y C W_y^\dagger D] &= \int d\mu(W) \text{Tr}[W A W^\dagger B W C W^\dagger D] \\ &= \frac{1}{d^2 - 1} (\text{Tr}[A] \text{Tr}[C] \text{Tr}[BD] + \text{Tr}[AC] \text{Tr}[B] \text{Tr}[D]) \\ &\quad - \frac{1}{d(d^2 - 1)} (\text{Tr}[AC] \text{Tr}[BD] + \text{Tr}[A] \text{Tr}[B] \text{Tr}[C] \text{Tr}[D]). \end{aligned} \quad (5)$$

Proof. The first equality follows from that fact that $\text{Tr}[W_y A W_y^\dagger B W_y C W_y^\dagger D] \in P_{(2,2)}(W_y)$. By writting

$$\text{Tr}[W A W^\dagger B W C W^\dagger D] = \sum_{\substack{\mathbf{i}_1, \mathbf{j}_1, \mathbf{i}'_1, \mathbf{j}'_1 \\ \mathbf{i}_2, \mathbf{j}_2, \mathbf{i}'_2, \mathbf{j}'_2}} a_{\mathbf{j}_1, \mathbf{j}'_1} b_{\mathbf{i}'_1, \mathbf{i}_2} c_{\mathbf{j}_2, \mathbf{j}'_2} d_{\mathbf{i}'_2, \mathbf{i}_1} w_{\mathbf{i}_1, \mathbf{j}_1} w_{\mathbf{i}_2, \mathbf{j}_2} w_{\mathbf{i}'_1, \mathbf{j}'_1}^* w_{\mathbf{i}'_2, \mathbf{j}'_2}^*,$$

we can use (3) to obtain

$$\begin{aligned} \int d\mu(W) \text{Tr}[W A W^\dagger B W C W^\dagger D] &= \frac{1}{d^2 - 1} \sum_{\mathbf{i}_1, \mathbf{j}_1, \mathbf{i}_2, \mathbf{j}_2} (a_{\mathbf{j}_1, \mathbf{j}_1} b_{\mathbf{i}_1, \mathbf{i}_2} c_{\mathbf{j}_2, \mathbf{j}_2} d_{\mathbf{i}_2, \mathbf{i}_1} + a_{\mathbf{j}_1, \mathbf{j}_2} b_{\mathbf{i}_2, \mathbf{i}_2} c_{\mathbf{j}_2, \mathbf{j}_1} d_{\mathbf{i}_1, \mathbf{i}_1}) \\ &\quad - \frac{1}{d(d^2 - 1)} \sum_{\mathbf{i}_1, \mathbf{j}_1, \mathbf{i}_2, \mathbf{j}_2} (a_{\mathbf{j}_1, \mathbf{j}_2} b_{\mathbf{i}_1, \mathbf{i}_2} c_{\mathbf{j}_2, \mathbf{j}_1} d_{\mathbf{i}_2, \mathbf{i}_1} + a_{\mathbf{j}_1, \mathbf{j}_1} b_{\mathbf{i}_2, \mathbf{i}_2} c_{\mathbf{j}_2, \mathbf{j}_2} d_{\mathbf{i}_1, \mathbf{i}_1}) \\ &= \frac{1}{d^2 - 1} (\text{Tr}[A] \text{Tr}[C] \text{Tr}[BD] + \text{Tr}[AC] \text{Tr}[B] \text{Tr}[D]) \\ &\quad - \frac{1}{d(d^2 - 1)} (\text{Tr}[AC] \text{Tr}[BD] + \text{Tr}[A] \text{Tr}[B] \text{Tr}[C] \text{Tr}[D]). \end{aligned} \quad (6)$$

□

Lemma 3. Let $\{W_y\}_{y \in Y} \subset U(d)$ form a unitary t -design with $t \geq 2$ and let $A, B, C, D : \mathcal{H}_w \rightarrow \mathcal{H}_w$ be arbitrary linear operators. Then

$$\begin{aligned} \frac{1}{|Y|} \sum_{y \in Y} \text{Tr}[W_y A W_y^\dagger B] \text{Tr}[W_y C W_y^\dagger D] &= \int d\mu(W) \text{Tr}[W A W^\dagger B] \text{Tr}[W C W^\dagger D] \\ &= \frac{1}{d^2 - 1} (\text{Tr}[A] \text{Tr}[B] \text{Tr}[C] \text{Tr}[D] + \text{Tr}[AC] \text{Tr}[BD]) \\ &\quad - \frac{1}{d(d^2 - 1)} (\text{Tr}[AC] \text{Tr}[B] \text{Tr}[D] + \text{Tr}[A] \text{Tr}[C] \text{Tr}[BD]). \end{aligned}$$

Proof. The first equality follows from a reasoning similar to the one used in Lemma 2. By expressing

$$\text{Tr}[W A W^\dagger B] \text{Tr}[W C W^\dagger D] = \sum_{\alpha, \beta} \text{Tr}[W A W^\dagger B |\alpha\rangle \langle \beta| W C W^\dagger D |\beta\rangle \langle \alpha|],$$

we can employ (5) to obtain

$$\begin{aligned}
\int d\mu(W) \text{Tr}[WAW^\dagger B] \text{Tr}[WCW^\dagger D] &= \sum_{\alpha, \beta} \int d\mu(W) \text{Tr}[WAW^\dagger B|\alpha\rangle\langle\beta|WCW^\dagger D|\beta\rangle\langle\alpha|] \\
&= \frac{1}{d^2 - 1} \sum_{\alpha, \beta} (\text{Tr}[A]\text{Tr}[C]\langle\alpha|B|\alpha\rangle\langle\beta|D|\beta\rangle + \text{Tr}[AC]\langle\beta|B|\alpha\rangle\langle\alpha|D|\beta\rangle) \\
&\quad - \frac{1}{d(d^2 - 1)} \sum_{\alpha, \beta} (\text{Tr}[AC]\langle\alpha|B|\alpha\rangle\langle\beta|D|\beta\rangle + \text{Tr}[A]\text{Tr}[C]\langle\beta|B|\alpha\rangle\langle\alpha|D|\beta\rangle) \\
&= \frac{1}{d^2 - 1} (\text{Tr}[A]\text{Tr}[B]\text{Tr}[C]\text{Tr}[D] + \text{Tr}[AC]\text{Tr}[BD]) \\
&\quad - \frac{1}{d(d^2 - 1)} (\text{Tr}[AC]\text{Tr}[B]\text{Tr}[D] + \text{Tr}[A]\text{Tr}[C]\text{Tr}[BD]) .
\end{aligned}$$

□

Lemma 4. Let $\mathcal{H} = \mathcal{H}_{\bar{w}} \otimes \mathcal{H}_w$ be a bipartite Hilbert space of dimension $d = d_{\bar{w}}d_w$, and let $\{W_y\}_{y \in Y}$ be a unitary t -design with $t \geq 1$ such that $W_y \in U(d_w)$ for all $y \in Y$. Then for arbitrary linear operators $A, B : \mathcal{H} \rightarrow \mathcal{H}$, we have

$$\int d\mu(W) (\mathbb{1}_{\bar{w}} \otimes W) A (\mathbb{1}_{\bar{w}} \otimes W^\dagger) B = \frac{\text{Tr}_w[A] \otimes \mathbb{1}_w}{d_w} B, \quad (7)$$

and

$$\int d\mu(W) \text{Tr}[(\mathbb{1}_{\bar{w}} \otimes W) A (\mathbb{1}_{\bar{w}} \otimes W^\dagger) B] = \frac{1}{d_w} \text{Tr}[\text{Tr}_w[A] \text{Tr}_w[B]]. \quad (8)$$

Here we use the notation $\mathbb{1}_w$ to indicate the identity operator on subsystem $\mathcal{H}_w$, and we employ Tr_w to indicate the partial trace over $\mathcal{H}_w$.

Proof. First, note that

$$(\mathbb{1}_{\bar{w}} \otimes W) A (\mathbb{1}_{\bar{w}} \otimes W^\dagger) B = \sum_{\mathbf{i}, \mathbf{j}, \mathbf{i}', \mathbf{j}'} w_{\mathbf{i}, \mathbf{j}} w_{\mathbf{i}', \mathbf{j}'}^* (\mathbb{1}_{\bar{w}} \otimes |\mathbf{i}\rangle\langle\mathbf{j}|) A (\mathbb{1}_{\bar{w}} \otimes |\mathbf{j}'\rangle\langle\mathbf{i}'|) B,$$

by using (2) we have

$$\begin{aligned}
\int d\mu(W) (\mathbb{1}_{\bar{w}} \otimes W) A (\mathbb{1}_{\bar{w}} \otimes W^\dagger) B &= \sum_{\mathbf{i}, \mathbf{j}, \mathbf{i}', \mathbf{j}'} \int d\mu(W) w_{\mathbf{i}, \mathbf{j}} w_{\mathbf{i}', \mathbf{j}'}^* (\mathbb{1}_{\bar{w}} \otimes |\mathbf{i}\rangle\langle\mathbf{j}|) A (\mathbb{1}_{\bar{w}} \otimes |\mathbf{j}'\rangle\langle\mathbf{i}'|) B \\
&= \frac{1}{d_w} \sum_{\mathbf{i}, \mathbf{j}} (\mathbb{1}_{\bar{w}} \otimes |\mathbf{i}\rangle\langle\mathbf{j}|) A (\mathbb{1}_{\bar{w}} \otimes |\mathbf{j}\rangle\langle\mathbf{i}|) B \\
&= \frac{1}{d_w} (\text{Tr}_w[A] \otimes \mathbb{1}_w) B.
\end{aligned}$$

Finally we can also obtain

$$\int d\mu(W) \text{Tr}[(\mathbb{1}_{\bar{w}} \otimes W) A (\mathbb{1}_{\bar{w}} \otimes W^\dagger) B] = \frac{1}{d_w} \text{Tr}[\text{Tr}_w[A] \text{Tr}_w[B]].$$

□

Lemma 5. If $\mathcal{H} = \mathcal{H}_{\bar{w}} \otimes \mathcal{H}_w$ is a bipartite Hilbert space of dimension $d = d_{\bar{w}}d_w$ ($d = 2^n$, and $\bar{d} = 2^{n'}$), and if $A, B : \mathcal{H} \rightarrow \mathcal{H}$ are arbitrary linear operators, then for any linear operator $W : \mathcal{H}_w \rightarrow \mathcal{H}_w$ we have

$$\text{Tr}[(\mathbb{1}_{\bar{w}} \otimes W) A (\mathbb{1}_{\bar{w}} \otimes W^\dagger) B] = \sum_{\mathbf{p}, \mathbf{q}} \text{Tr}[W A_{\mathbf{qp}} W^\dagger B_{\mathbf{pq}}], \quad (9)$$

where the summation runs over all bitstrings of length n' , and where

$$A_{\mathbf{qp}} = \text{Tr}_{\bar{w}}[(|\mathbf{p}\rangle\langle\mathbf{q}| \otimes \mathbb{1}_w) A], \quad B_{\mathbf{pq}} = \text{Tr}_{\bar{w}}[(|\mathbf{q}\rangle\langle\mathbf{p}| \otimes \mathbb{1}_w) B]. \quad (10)$$

Proof. By expanding the operators in the computational basis

$$\mathbb{1}_{\bar{w}} \otimes W = \sum_{\mathbf{p}, \mathbf{i}, \mathbf{j}} w_{\mathbf{i}, \mathbf{j}} |\mathbf{p}\rangle \langle \mathbf{p}| \otimes |\mathbf{i}\rangle \langle \mathbf{j}|, \quad \mathbb{1}_{\bar{w}} \otimes W^\dagger = \sum_{\mathbf{q}, \mathbf{i}', \mathbf{j}'} w_{\mathbf{i}', \mathbf{j}'}^* |\mathbf{q}\rangle \langle \mathbf{q}| \otimes |\mathbf{j}'\rangle \langle \mathbf{i}'|, \quad (11)$$

$$A = \sum_{\substack{\mathbf{k}_1, \mathbf{k}_2 \\ \mathbf{l}_1, \mathbf{l}_2}} a_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{l}_1, \mathbf{l}_2} |\mathbf{k}_1\rangle \langle \mathbf{l}_1| \otimes |\mathbf{k}_2\rangle \langle \mathbf{l}_2|, \quad B = \sum_{\substack{\mathbf{p}_1, \mathbf{p}_2 \\ \mathbf{q}_1, \mathbf{q}_2}} b_{\mathbf{p}_1, \mathbf{p}_2, \mathbf{q}_1, \mathbf{q}_2} |\mathbf{p}_1\rangle \langle \mathbf{q}_1| \otimes |\mathbf{p}_2\rangle \langle \mathbf{q}_2|, \quad (12)$$

we have

$$\text{Tr}[(\mathbb{1}_{\bar{w}} \otimes W)A(\mathbb{1}_{\bar{w}} \otimes W^\dagger)B] = \sum_{\substack{\mathbf{i}, \mathbf{j}, \mathbf{i}', \mathbf{j}' \\ \mathbf{p}, \mathbf{q}}} w_{\mathbf{i}, \mathbf{j}} a_{\mathbf{p}, \mathbf{j}, \mathbf{q}, \mathbf{i}'} w_{\mathbf{i}', \mathbf{j}'}^* b_{\mathbf{q}, \mathbf{j}', \mathbf{p}, \mathbf{i}} = \sum_{\mathbf{p}, \mathbf{q}} \text{Tr}[W A_{\mathbf{qp}} W^\dagger B_{\mathbf{pq}}]. \quad (13)$$

□

Lemma 6. Let $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3 \otimes \mathcal{H}_4$ be a Hilbert space of dimension $d \cdot d \cdot d \cdot d = d^4$, and let $W : \mathcal{H}_1 \otimes \mathcal{H}_2 \rightarrow \mathcal{H}_1 \otimes \mathcal{H}_2$ be a t -design with $t \geq 2$. For arbitrary linear operators $A, A' : \mathcal{H} \rightarrow \mathcal{H}$ we define

$$\Omega_1 = (W \otimes \mathbb{1}_3 \otimes \mathbb{1}_4)A(W^\dagger \otimes \mathbb{1}_3 \otimes \mathbb{1}_4)(|\mathbf{p}\rangle \langle \mathbf{q}| \otimes \mathbb{1}_2 \otimes \mathbb{1}_3 \otimes \mathbb{1}_4), \quad \Omega_2 = (W \otimes \mathbb{1}_3 \otimes \mathbb{1}_4)A(W^\dagger \otimes \mathbb{1}_3 \otimes \mathbb{1}_4), \quad (14)$$

$$\Omega'_1 = (W \otimes \mathbb{1}_3 \otimes \mathbb{1}_4)A'(W^\dagger \otimes \mathbb{1}_3 \otimes \mathbb{1}_4)(|\mathbf{p}'\rangle \langle \mathbf{q}'| \otimes \mathbb{1}_2 \otimes \mathbb{1}_3 \otimes \mathbb{1}_4), \quad \Omega'_2 = (W \otimes \mathbb{1}_3 \otimes \mathbb{1}_4)A'(W^\dagger \otimes \mathbb{1}_3 \otimes \mathbb{1}_4), \quad (15)$$

where $\mathbf{p}, \mathbf{q}, \mathbf{p}'$, and $\mathbf{q}'$ are bitstrings of length d .

Let us first consider the quantity

$$\Delta\Omega_{jkl}^{(1)} = \text{Tr}_{kl}[\text{Tr}_{1j}[\Omega_1]\text{Tr}_{1j}[\Omega'_1]] - \frac{\text{Tr}_L[\text{Tr}_{1jk}[\Omega_1]_{1jk} \text{Tr}[\Omega'_1]]}{d_k}, \quad (16)$$

where Tr_j indicates the partial trace over $\mathcal{H}_j$, and where $\text{Tr}_{jk} := \text{Tr}_j \text{Tr}_k$. Here the indexes j, k , and l define Hilbert spaces such that $\mathcal{H}_j \otimes \mathcal{H}_k \otimes \mathcal{H}_L = \mathcal{H}_2 \otimes \mathcal{H}_3 \otimes \mathcal{H}_4$. In addition, let us define $\mathcal{H}_j = \mathcal{H}_0 := \{\mathbf{0}\}$, in which case for any operator O we also define the partial trace over $\mathcal{H}_0$ as $\text{Tr}_0[O] := O$, and $\text{Tr}_{k0}[O] := \text{Tr}_k[O]$. From the previous definition, we then have that $\mathcal{H}_k$ can be chosen from the set $\{\mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_2 \otimes \mathcal{H}_3, \mathcal{H}_3 \otimes \mathcal{H}_4\}$, and for any such choice the following equality always holds

$$\begin{aligned} \int d\mu(W) \Delta\Omega_{jkl}^{(1)} &= \frac{t_1^{(1)} \delta_{\mathbf{pq}'} \delta_{\mathbf{p}'\mathbf{q}}}{d^4 - 1} \left(\text{Tr}_{123x}[\text{Tr}_y[A]\text{Tr}_y[A']] - \frac{\text{Tr}_{3x}[\text{Tr}_{12y}[A]\text{Tr}_{12y}[A']]}{d^2} \right) \\ &+ \frac{t_2^{(1)} \delta_{\mathbf{pq}'} \delta_{\mathbf{p}'\mathbf{q}'}}{d^4 - 1} \left(\text{Tr}_{3x}[\text{Tr}_{12y}[A]\text{Tr}_{12y}[A']] - \frac{\text{Tr}_x[\text{Tr}_{123y}[A]\text{Tr}_{123y}[A']]}{d'} \right) \\ &- \frac{t_3^{(1)} \delta_{\mathbf{pq}'} \delta_{\mathbf{p}'\mathbf{q}}}{d^4 - 1} \left(\text{Tr}_{12x}[\text{Tr}_{3y}[A]\text{Tr}_{3y}[A']] - \frac{\text{Tr}_x[\text{Tr}_{123y}[A]\text{Tr}_{123y}[A']]}{d^2} \right) \\ &- \frac{t_4^{(1)} \delta_{\mathbf{pq}'} \delta_{\mathbf{p}'\mathbf{q}'}}{d^4 - 1} \left(\text{Tr}_{123x}[\text{Tr}_y[A]\text{Tr}_y[A']] - \frac{\text{Tr}_{12x}[\text{Tr}_{3y}[A]\text{Tr}_{3y}[A']]}{d'} \right) \end{aligned} \quad (17)$$

where $d' = d^2$ if $\mathcal{H}_4 \leq \mathcal{H}_k$, and $d' = d$ otherwise. Here we employ the notation $A \leq B$ to indicate that A is a subspace of B . Note that $\mathcal{H}_y = \mathcal{H}_4$, $\mathcal{H}_x = \mathcal{H}_0$ if $\mathcal{H}_4 \leq \mathcal{H}_j$, and $\mathcal{H}_y = \mathcal{H}_0$, $\mathcal{H}_x = \mathcal{H}_4$ otherwise. In addition, we have

$$t_1^{(1)} = \begin{cases} d & \text{if } \mathcal{H}_2 \leq \mathcal{H}_j \\ 0 & \text{if } \mathcal{H}_k = \mathcal{H}_2 \\ d^2 & \text{otherwise} \end{cases}, \quad t_2^{(1)} = \begin{cases} d^2 & \text{if } \mathcal{H}_2 \leq \mathcal{H}_j \\ 0 & \text{if } \mathcal{H}_k = \mathcal{H}_2 \\ d & \text{otherwise} \end{cases}, \quad t_4^{(1)} = \begin{cases} 1 & \text{if } \mathcal{H}_2 \leq \mathcal{H}_j, \text{ and } \mathcal{H}_3 \in \mathcal{H}_k \\ 0 & \text{if } \mathcal{H}_k = \mathcal{H}_1 \\ \frac{1}{d} & \text{otherwise} \end{cases} \quad (18)$$

$$t_3^{(1)} = \begin{cases} d & \text{if } \mathcal{H}_k = \mathcal{H}_3, \text{ and } \mathcal{H}_2 \leq \mathcal{H}_L \\ 1 - d^2 & \text{if } \mathcal{H}_k = \mathcal{H}_2 \\ \frac{1}{d} & \text{if } \mathcal{H}_k = \mathcal{H}_2 \otimes \mathcal{H}_3, \text{ or if } \mathcal{H}_k = \mathcal{H}_3 \otimes \mathcal{H}_4, \text{ and } \mathcal{H}_j = \mathcal{H}_2 \\ 1 & \text{otherwise} \end{cases}, \quad (19)$$

where we can verify that for all 12 possible cases that $\sum_{k=1}^4 |\frac{t_k^{(1)}}{(d^4-1)}| \leq 1$.

Second, let us consider the quantity

$$\Delta\Omega_{jkl}^{(2)} = \text{Tr}_{kl}[\text{Tr}_{1j}[\Omega_2]\text{Tr}_{1j}[\Omega'_2]] - \frac{\text{Tr}_L[\text{Tr}_{12j}[\Omega_2]_{1jk} \text{Tr}[\Omega'_2]]}{d_k}. \quad (20)$$

Then, for $\mathcal{H}_k \in \{\mathcal{H}_2, \mathcal{H}_3, \mathcal{H}_2 \otimes \mathcal{H}_3, \mathcal{H}_3 \otimes \mathcal{H}_4\}$, the following equality always holds

$$\begin{aligned} \int d\mu(W) \Delta\Omega_{jkl}^{(2)} = & \frac{t_1^{(2)}}{d^2 + 1} \left(\text{Tr}_{123x}[\text{Tr}_y[A]\text{Tr}_y[A']] - \frac{\text{Tr}_{3x}[\text{Tr}_{12y}[A]\text{Tr}_{12y}[A']]}{d^2} \right) \\ & + \frac{t_2^{(2)}}{d^2 + 1} \left(\text{Tr}_{3x}[\text{Tr}_{12y}[A]\text{Tr}_{12y}[A']] - \frac{\text{Tr}_x[\text{Tr}_{123y}[A]\text{Tr}_{123y}[A']]}{d'} \right) \\ & - \frac{t_3^{(2)}}{d^2 + 1} \left(\text{Tr}_{12x}[\text{Tr}_{3y}[A]\text{Tr}_{3y}[A']] - \frac{\text{Tr}_x[\text{Tr}_{123y}[A]\text{Tr}_{123y}[A']]}{d^2} \right), \end{aligned} \quad (21)$$

where $d' = d^2$ if $\mathcal{H}_4 \leq \mathcal{H}_k$, and $d' = d$ otherwise. Similarly, $\mathcal{H}_y = \mathcal{H}_4$, $\mathcal{H}_x = \mathcal{H}_0$ if $\mathcal{H}_4 \leq \mathcal{H}_j$, and $\mathcal{H}_y = \mathcal{H}_0$, $\mathcal{H}_x = \mathcal{H}_4$ otherwise. In addition, we now have

$$t_1^{(2)} = \begin{cases} 0 & \text{if } \mathcal{H}_2 \leq \mathcal{H}_j \\ d & \text{otherwise} \end{cases}, \quad t_2^{(2)} = \begin{cases} 0 & \text{if } \mathcal{H}_2 \leq \mathcal{H}_j, \text{ or } \mathcal{H}_k = \mathcal{H}_2 \\ \frac{d^2+1}{d} & \text{otherwise} \end{cases}, \quad t_3^{(2)} = \begin{cases} 1 & \text{if } \mathcal{H}_2 \leq \mathcal{H}_L, \text{ and } \mathcal{H}_k = \mathcal{H}_3 \\ \frac{1}{d} & \text{if } \mathcal{H}_L = \mathcal{H}_2, \text{ and } \mathcal{H}_k = \mathcal{H}_3 \otimes \mathcal{H}_4 \\ 0 & \text{otherwise} \end{cases}, \quad (22)$$

where we can verify that for all 9 nontrivial cases $\sum_{k=1}^3 | \frac{t_k^{(2)}}{d^2+1} | \leq 1$.

Proof. The proof of Eqs. (17), and (21) can be obtain by explicitly integrating each term via (3). $\square$

In particular, let us consider from Eq. (17) the following case which will be relevant for our proofs.

$$\begin{aligned} \int d\mu(W) \left(\text{Tr}_{34} [\text{Tr}_{12}[\Omega_1]\text{Tr}_{12}[\Omega'_1]] - \frac{\text{Tr}[\Omega_1]\text{Tr}[\Omega'_1]}{d^2} \right) = & \frac{d^2 \delta_{\mathbf{p}_2 \mathbf{q}'_2} \delta_{\mathbf{p}'_2 \mathbf{q}_2}}{d^4 - 1} \left(\text{Tr}_{234}[\text{Tr}_1[A]\text{Tr}_1[A']] - \frac{\text{Tr}_4[\text{Tr}_{123}[A]\text{Tr}_{123}[A']]}{d^2} \right) \\ & + \frac{d \delta_{\mathbf{p}_2 \mathbf{q}_2} \delta_{\mathbf{p}'_2 \mathbf{q}'_2}}{d^4 - 1} \left(\text{Tr}_4[\text{Tr}_{123}[A]\text{Tr}_{123}[A']] - \frac{\text{Tr}[A]\text{Tr}[A']}{d} \right) \\ & - \frac{\delta_{\mathbf{p}_2 \mathbf{q}'_2} \delta_{\mathbf{p}'_2 \mathbf{q}_2}}{d(d^4 - 1)} \left(\text{Tr}_{23}[\text{Tr}_{14}[A]\text{Tr}_{14}[A']] - \frac{\text{Tr}[A]\text{Tr}[A']}{d^2} \right) \\ & - \frac{\delta_{\mathbf{p}_2 \mathbf{q}_2} \delta_{\mathbf{p}'_2 \mathbf{q}'_2}}{d(d^4 - 1)} \left(\text{Tr}_{234}[\text{Tr}_1[A]\text{Tr}_1[A']] - \frac{\text{Tr}_{23}[\text{Tr}_{14}[A]\text{Tr}_{14}[A']]}{d} \right). \end{aligned} \quad (23)$$

Lemma 7. Let $\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2 \otimes \mathcal{H}_3$ be a tripartite Hilbert space of dimension $d = d_1 d_2 d_3$, and let $O_1 = O_A \otimes \mathbb{1}_2 \otimes \mathbb{1}_3$, and $O_2 = \mathbb{1}_1 \otimes O_B \otimes \mathbb{1}_3$ be linear operators on S . Here $\mathbb{1}_i$ indicates the identity over subsystem S_i , so that O_1 and O_2 have no overlapping support. Then for any linear operators $O_A : S_1 \rightarrow S_1$, and $O_B : S_2 \rightarrow S_2$ we have

$$\text{Tr}_{jk} [\text{Tr}_i [O_1] \text{Tr}_i [O_2]] - \frac{\text{Tr}_k [\text{Tr}_{ij} [O_1] \text{Tr}_{ij} [O_2]]}{d_j} = 0, \quad (24)$$

where Tr_i indicates the partial trace over $\mathcal{H}_i$, $\text{Tr}_{ij} := \text{Tr}_i \text{Tr}_j$, and where we defined $\mathcal{H}_i \otimes \mathcal{H}_j \otimes \mathcal{H}_k = \mathcal{H}$ such that $\mathcal{H}_j = \mathcal{H}_1, \mathcal{H}_2$ or $\mathcal{H}_3$. Moreover, one can always choose $\mathcal{H}_i = \mathcal{H}_0 := \{\mathbf{0}\}$ (or $\mathcal{H}_k = \mathcal{H}_0$), in which case we define the partial trace over $\mathcal{H}_0$ as $\text{Tr}_0 [O] := O$, and $\text{Tr}_{j0} [O] := \text{Tr}_j [O]$.

Proof. Let us show how this equality holds for the specific case when $\mathcal{H}_i = \{\mathbf{0}\}$, $\mathcal{H}_j = \mathcal{H}_2$, and $\mathcal{H}_k = \mathcal{H}_1 \otimes \mathcal{H}_3$, and let us remark that all remaining cases follow similarly. We have

$$\text{Tr}_0 [O_1] \text{Tr}_0 [O_2] = O_A \otimes O_B \otimes \mathbb{1}_3, \quad \text{Tr}_2 [O_1] = d_2 O_A \otimes \mathbb{1}_3, \quad \text{Tr}_2 [O_2] = \text{Tr}_2 [O_B] \mathbb{1}_1 \otimes \mathbb{1}_3, \quad (25)$$

and Eq. (24) becomes

$$\text{Tr}_{123} [O_1 O_2] - \frac{\text{Tr}_{13} [\text{Tr}_2 [O_1] \text{Tr}_2 [O_2]]}{d_2} = \text{Tr}_{123} [O_A \otimes O_B \otimes \mathbb{1}_3] - \frac{d_2 \text{Tr}_2 [O_B] \text{Tr}_{13} [O_A \otimes \mathbb{1}_3]}{d_2} \quad (26)$$

$$= d_3 \text{Tr}_1 [O_A] \text{Tr}_2 [O_B] - \frac{d_2 d_3 \text{Tr}_1 [O_A] \text{Tr}_2 [O_B]}{d_2} \quad (27)$$

$$= 0. \quad (28)$$

$\square$

Supplementary Note 2: Proof of Proposition 1

Here we provide a proof for Proposition 1, which we recall for convenience:

Proposition 1. Let θ^j be uniformly distributed on $[-\pi, \pi] \forall j$. For any $\delta \in (0, 1)$, the probability that $C_G \leq \delta$ satisfies

$$\Pr\{C_G \leq \delta\} \leq (1 - \delta)^{-1} \left(\frac{1}{2}\right)^n. \quad (29)$$

For any $\delta \in [\frac{1}{2}, 1]$, the probability that $C_L \leq \delta$ satisfies

$$\Pr\{C_L \leq \delta\} \geq \frac{(2\delta - 1)^2}{\frac{1}{2n} + (2\delta - 1)^2} \xrightarrow{n \rightarrow \infty} 1. \quad (30)$$

Proposition 1 formalizes the narrow gorge phenomenon shown in Fig. 2 of the main text for the warm-up example. Specifically, it bounds the volume of parameter space that is trainable in the sense that the cost deviates from its maximum value of one. As a consequence, one finds that the probability that the global cost function is less than 1 vanishes exponentially as $n \rightarrow \infty$, whereas for the local cost function there are neighborhoods of the minimum that have constant probability as $n \rightarrow \infty$. In the following proof, it is helpful to define the trainable region for the global and local cost as $A_\delta^G := \{\theta : C_G(\theta) \leq \delta\}$ and $A_\delta^L := \{\theta : C_L(\theta) \leq \delta\}$.

Proof. Let us first consider the global cost function

$$C_G = \text{Tr}[O_G V(\theta) |0\rangle\langle 0| V(\theta)^\dagger], \quad (31)$$

with $O_G = \mathbb{1} - |0\rangle\langle 0|$, and with $V(\theta) = \bigotimes_{j=1}^n e^{-i\theta^j \sigma_x^{(j)}/2}$. The probability of the region A_δ^G can be bounded from above by defining the random variable $X = \prod_{j=1}^n \cos^2 \frac{\theta_j}{2} = 1 - C_G$ and noting that $(1 - \delta)P\{X \geq 1 - \delta\} \leq E(X)$. Then

$$\begin{aligned} \Pr(A_\delta^G) &= P\left\{\prod_{j=1}^n \cos^2 \frac{\theta_j}{2} \geq 1 - \delta\right\} \\ &\leq (1 - \delta)^{-1} E\left(\prod_{j=1}^n \cos^2 \frac{\theta_j}{2}\right) \\ &= (1 - \delta)^{-1} \left(\frac{1}{2}\right)^n \end{aligned} \quad (32)$$

Therefore, for any $\delta \in (0, 1)$, the probability that $C_G \leq \delta$ goes to zero exponentially with n .

Now consider the local cost function C_L given by

$$C_L = \text{Tr}[O_L V(\theta) |0\rangle\langle 0| V(\theta)^\dagger], \quad \text{with } O_L = \mathbb{1} - \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}}, \quad (33)$$

and where $\mathbb{1}_{\bar{j}}$ is the identity on all qubits except qubit j . To show the dependence of the result on the range of local cost function values, we write a parametrized local cost function $C_L(\cdot; \lambda) := 1 - \frac{1}{\lambda n} \sum_{j=1}^n \cos^2 \frac{\theta_j}{2}$ and define $A_\delta^L := \{\theta : C_L(\theta; \lambda) \leq \delta\}$. C_L in the main text is obtained for $\lambda = 1$. The parameter λ is introduced to show that the range of validity of the inequality is dependent on the cost function. For δ in the interval $\delta \in (1 - (2\lambda)^{-1}, 1]$,

$$\Pr(A_\delta^L) = \Pr\left(\left\{\theta : \frac{1}{\lambda n} \sum_{j=1}^n \cos^2 \frac{\theta_j}{2} \geq 1 - \delta\right\}\right) \geq \frac{(1 - 2\lambda(1 - \delta))^2}{\frac{1}{2n} + (1 - 2\lambda(1 - \delta))^2} \quad (34)$$

which tends to 1 as $n \rightarrow \infty$. The inequality follows from the Paley-Zygmund inequality [4] in the form

$$P(X \geq rE(X)) \geq \frac{(1 - r)^2 E(X)^2}{\text{Var}X + (1 - r)^2 E(X)^2} \quad (35)$$

which holds for random variables $X \geq 0$ and scalar r with $0 \leq r \leq 1$. In particular, (34) follows by taking $X = \frac{1}{n\lambda} \sum_{j=1}^n \cos^2 \frac{\theta_j}{2}$ and $r = 2\lambda(1 - \delta)$, so that r varies from 1 to 0 as δ varies from $1 - (2\lambda)^{-1}$ to 1. $\square$

Supplementary Note 3: Proof of Proposition 2

Let us first recall that in the main text we analyze cost functions C which can be expressed as the expectation value of a given observable O as

$$C = \text{Tr} [OV(\boldsymbol{\theta})\rho V^\dagger(\boldsymbol{\theta})] , \quad (36)$$

where $V(\boldsymbol{\theta})$ is a parametrized quantum gate sequence and ρ is a general input mixed quantum state on n qubits. Here we provide a proof for Proposition 2, which we recall for convenience:

Proposition 2. The average of the partial derivative of any cost function of the form (36) with respect to a parameter θ^ν in a block W of the ansatz $V(\boldsymbol{\theta})$ is

$$\langle \partial_\nu C \rangle_V = 0 , \quad (37)$$

provided that either W_A or W_B form a 1-design.

As discussed in the main text, we employ an Alternating Layered Ansatz, where each layer is composed of m -qubits gates or “blocks”. In particular, each block $W_{kl}(\boldsymbol{\theta}_{kl})$ in $V(\boldsymbol{\theta})$ can be written as a product of ζ_{kl} independent gates from a gate alphabet $\mathcal{A} = \{G_\mu(\theta)\}$ as

$$W_{kl}(\boldsymbol{\theta}_{kl}) = G_{\zeta_{kl}}(\theta_{kl}^{\zeta_{kl}}) \dots G_\nu(\theta_{kl}^\nu) \dots G_1(\theta_{kl}^1) , \quad (38)$$

where θ_{kl}^ν are continuous parameter, and where $G_\nu(\theta_{kl}^\nu) = R_\nu(\theta_{kl}^\nu)Q_\nu$ with Q_ν an unparametrized gate, and $R_\nu(\theta_{kl}^\nu) = e^{-i\theta_{kl}^\nu \sigma_\nu/2}$ such that σ_ν is a Pauli operator.

Consider now a block $W_{kl}(\boldsymbol{\theta}_{kl})$ in the l -th layer of the ansatz. For the rest of this Supplementary Information we simply use the notation W when referring to this particular block. First, let S_w denote the m -qubit subsystem that contains the qubits W acts on, and let $S_{\bar{w}}$ be the $(n - m)$ subsystem on which W acts trivially, with $\mathcal{H}_w$, and $\mathcal{H}_{\bar{w}}$ their respective associated Hilbert spaces. Then, consider a given trainable parameter θ^ν in W , such that we can express $W = W_B W_A$, with

$$W_B = \prod_{\mu=1}^{\nu-1} G_\mu(\theta^\mu) , \quad \text{and} \quad W_A = \prod_{\mu=\nu}^{\zeta} G_\mu(\theta^\mu) . \quad (39)$$

Then, we recall that we have defined the forward light-cone $\mathcal{L}$ of W as all gates with at least one input qubit causally connected to the output qubits of W . We can then define $S_{\mathcal{L}}$ as the subsystem of all qubits in $\mathcal{L}$, and $\mathcal{H}_{\mathcal{L}}$ as its associated Hilbert space. Without loss of generality, the trainable gate sequence can be expressed as

$$V(\boldsymbol{\theta}) = V_R(\mathbb{1}_{\bar{w}} \otimes W) V_L , \quad (40)$$

where $\mathbb{1}_{\bar{w}}$ indicates the identity in $\mathcal{H}_{\bar{w}}$, and where we assume without loss of generality that V_R contains the gates in $\mathcal{L}$ and all the blocks W_{kL} in the last layer of $V(\boldsymbol{\theta})$.

Proof. The partial derivative of W with respect to the angle θ_ν is given by

$$\partial_\nu W = W_A \left(-\frac{i}{2} \sigma_\nu \right) W_B , \quad (41)$$

where here σ_ν is an operator $\sigma_\nu : \mathcal{H}_w \rightarrow \mathcal{H}_w$, which acts non-trivially on a qubit given qubit j in $\mathcal{H}_w$, i.e., $\sigma_\nu := (\sigma_\nu)_j \otimes \mathbb{1}_{\bar{w}}$. Hence, by means of Eqs. (41) and (40) we have

$$\begin{aligned} \partial_\nu C &= \text{Tr} [O (\partial_\nu V(\boldsymbol{\theta})) \rho V^\dagger(\boldsymbol{\theta}) + V(\boldsymbol{\theta}) \rho (\partial_\nu V^\dagger(\boldsymbol{\theta}))] \\ &= \text{Tr} \left[OV_R(\mathbb{1}_{\bar{w}} \otimes W_A) \left(\mathbb{1}_{\bar{w}} \otimes \left(-\frac{i}{2} \sigma_\nu \right) \right) (\mathbb{1}_{\bar{w}} \otimes W_B) V_L \rho V_L^\dagger \left(\mathbb{1}_{\bar{w}} \otimes W_B^\dagger W_A^\dagger \right) V_R^\dagger \right] \\ &\quad + \text{Tr} \left[OV_R(\mathbb{1}_{\bar{w}} \otimes W_A W_B) V_L \rho V_L^\dagger \left(\mathbb{1}_{\bar{w}} \otimes W_B^\dagger \right) \left(\mathbb{1}_{\bar{w}} \otimes \left(+\frac{i}{2} \sigma_\nu \right) \right) \left(\mathbb{1}_{\bar{w}} \otimes W_A^\dagger \right) V_R^\dagger \right] . \end{aligned}$$

Which can be simplified as

$$\partial_\nu C = \frac{i}{2} \text{Tr} \left[(\mathbb{1}_{\bar{w}} \otimes W_B) V_L \rho V_L^\dagger \left(\mathbb{1}_{\bar{w}} \otimes W_B^\dagger \right) \left[\mathbb{1}_{\bar{w}} \otimes \sigma_\nu, \left(\mathbb{1}_{\bar{w}} \otimes W_A^\dagger \right) V_R^\dagger OV_R(\mathbb{1}_{\bar{w}} \otimes W_A) \right] \right] , \quad (42)$$

or equivalently, as

$$\partial_\nu C = -\frac{i}{2} \text{Tr} \left[\left(\mathbb{1}_{\bar{w}} \otimes W_A^\dagger \right) V_R^\dagger O V_R \left(\mathbb{1}_{\bar{w}} \otimes W_A \right) \left[\mathbb{1}_{\bar{w}} \otimes \sigma_\nu, \left(\mathbb{1}_{\bar{w}} \otimes W_B \right) V_L \rho V_L^\dagger \left(\mathbb{1}_{\bar{w}} \otimes W_B^\dagger \right) \right] \right]. \quad (43)$$

In order to compute the expectation value $\langle \partial_\nu C \rangle_V$ we need to consider three different scenarios: (1) when only W_A is a 1-design; (2) when only W_B is a 1-design; and (3) when both W_A and W_B form 1-designs. We first consider the case when W_A is a 1-design. Since W_A , W_B , V_R and V_L are independent, we can compute the expectation value over the ansatz as $\langle \partial_\nu C \rangle_V = \langle \langle \partial_\nu C \rangle_{W_A} \rangle_{V_L, W_B, V_R}$. From (42) and the definition of a 1-design in (1), we can compute

$$\begin{aligned} \langle \partial_\nu C \rangle_{W_A} &= -\frac{i}{2} \text{Tr} \left[\left(\mathbb{1}_{\bar{w}} \otimes W_B \right) V_L \rho V_L^\dagger \left(\mathbb{1}_{\bar{w}} \otimes W_B^\dagger \right) \left[\mathbb{1}_{\bar{w}} \otimes \sigma_\nu, \int d\mu(W_A) \left(\mathbb{1}_{\bar{w}} \otimes W_A \right) V_R^\dagger O V_R \left(\mathbb{1}_{\bar{w}} \otimes W_A^\dagger \right) \right] \right] \\ &= -\frac{i}{2} \text{Tr} \left[\left(\mathbb{1}_{\bar{w}} \otimes W_B \right) V_L \rho V_L^\dagger \left(\mathbb{1}_{\bar{w}} \otimes W_B^\dagger \right) \left[\mathbb{1}_{\bar{w}} \otimes \sigma_\nu, \frac{1}{2^m} \text{Tr}_w [V_R^\dagger O V_R] \otimes \mathbb{1}_w \right] \right] \\ &= 0, \end{aligned} \quad (44)$$

where in the second equality we used Lemma 4.

On the other hand, if W_B is a 1-design, we can now employ (43) to get

$$\begin{aligned} \langle \partial_\nu C \rangle_{W_B} &= -\frac{i}{2} \text{Tr} \left[\left(\mathbb{1}_{\bar{w}} \otimes W_A^\dagger \right) V_R^\dagger O V_R \left(\mathbb{1}_{\bar{w}} \otimes W_A \right) \left[\mathbb{1}_{\bar{w}} \otimes \sigma_\nu, \int d\mu(W_B) \left(\mathbb{1}_{\bar{w}} \otimes W_B \right) V_L \rho V_L^\dagger \left(\mathbb{1}_{\bar{w}} \otimes W_B^\dagger \right) \right] \right] \\ &= -\frac{i}{2} \text{Tr} \left[\left(\mathbb{1}_{\bar{w}} \otimes W_A^\dagger \right) V_R^\dagger O V_R \left(\mathbb{1}_{\bar{w}} \otimes W_A \right) \left[\mathbb{1}_{\bar{w}} \otimes \sigma_\nu, \frac{1}{2^m} \text{Tr}_w [V_L \rho V_L^\dagger] \otimes \mathbb{1}_w \right] \right] \\ &= 0, \end{aligned} \quad (45)$$

which follows from the same argument used to derive (44). Finally, from Eqs. (44) and (45), we have $\langle \partial_\nu C \rangle_V = 0$ ($\forall V_L, V_R$) if W_A and W_B are both 1-designs. $\square$

Supplementary Note 4: Variance of the cost function partial derivative

In this section, we derive the formula for the variance of the cost function gradient. Namely,

$$\langle (\partial_\nu C)^2 \rangle_V = \frac{2^{m-1} \text{Tr}[\sigma_\nu^2]}{(2^{2m} - 1)^2} \sum_{\substack{pq \\ p'q'}} \left\langle \Delta \Omega_{pq}^{p'q'} \right\rangle_{V_R} \left\langle \Delta \Psi_{pq}^{p'q'} \right\rangle_{V_L}, \quad (46)$$

with

$$\Delta \Omega_{pq}^{p'q'} = \text{Tr}[\Omega_{qp} \Omega_{q'p'}] - \frac{\text{Tr}[\Omega_{qp}] \text{Tr}[\Omega_{q'p'}]}{2^m}, \quad (47)$$

$$\Delta \Psi_{pq}^{p'q'} = \text{Tr}[\Psi_{pq} \Psi_{p'q'}] - \frac{\text{Tr}[\Psi_{pq}] \text{Tr}[\Psi_{p'q'}]}{2^m}, \quad (48)$$

and where

$$\Omega_{qp} = \text{Tr}_{\bar{w}} \left[(|p\rangle \langle q| \otimes \mathbb{1}_w) V_R^\dagger O V_R \right], \quad (49)$$

$$\Psi_{pq} = \text{Tr}_{\bar{w}} \left[(|q\rangle \langle p| \otimes \mathbb{1}_w) V_L \rho V_L^\dagger \right]. \quad (50)$$

Proof. As shown in the previous section, $\langle \partial_\nu C \rangle_V = 0$ when either W_B or W_A of (39) are 1-designs. Hence, we can compute the variance of $\partial_\nu C$ as $\text{Var}[\partial_\nu C] = \langle (\partial_\nu C)^2 \rangle_V - \langle \partial_\nu C \rangle_V^2 = \langle (\partial_\nu C)^2 \rangle_V$. From (43) and Lemma 5 we have

$$(\partial_\nu C)^2 = -\frac{1}{4} \sum_{\substack{p,q \\ p',q'}} \text{Tr} \left[W_A \Omega_{qp} W_A^\dagger \Gamma_{pq} \right] \text{Tr} \left[W_A \Omega_{q'p'} W_A^\dagger \Gamma_{p'q'} \right], \quad (51)$$

with

$$\begin{aligned}
\Gamma_{\mathbf{pq}} &= \text{Tr}_{\bar{w}} \left[(|\mathbf{q}\rangle\langle\mathbf{p}| \otimes \mathbb{1}_w) \left[\mathbb{1}_{\bar{w}} \otimes \sigma_\nu, (\mathbb{1}_{\bar{w}} \otimes W_B) V_L \rho V_L^\dagger (\mathbb{1}_{\bar{w}} \otimes W_B^\dagger) \right] \right] \\
&= \text{Tr}_{\bar{w}} \left[[\mathbb{1}_{\bar{w}} \otimes \sigma_\nu, (\mathbb{1}_{\bar{w}} \otimes W_B) (|\mathbf{q}\rangle\langle\mathbf{p}| \otimes \mathbb{1}_w) V_L \rho V_L^\dagger (\mathbb{1}_{\bar{w}} \otimes W_B^\dagger)] \right] \\
&= [\sigma_\nu, W_B \text{Tr}_{\bar{w}}[(|\mathbf{q}\rangle\langle\mathbf{p}| \otimes \mathbb{1}_w) V_L \rho V_L^\dagger] W_B^\dagger] \\
&= [\sigma_\nu, W_B \Psi_{\mathbf{pq}} W_B^\dagger] .
\end{aligned} \tag{52}$$

As previously mentioned, if W_A , W_B , V_R and V_L are independent, the expectation value of (51) can be computed as $\langle(\partial_\nu C)^2\rangle_V = \langle(\partial_\nu C)^2\rangle_{V_L, W_B, W_A, V_R}$. In addition, if W_A is a 2-design, we get from Lemma 3

$$\begin{aligned}
\langle(\partial_\nu C)^2\rangle_{W_A} &= -\frac{1}{4} \sum_{\substack{\mathbf{p}, \mathbf{q} \\ \mathbf{p}', \mathbf{q}'}} \int d\mu(W_A) \text{Tr} [W_A \Omega_{\mathbf{qp}} W_A^\dagger \Gamma_{\mathbf{pq}}] \text{Tr} [W_A \Omega_{\mathbf{q'p'}} W_A^\dagger \Gamma_{\mathbf{p'q'}}] \\
&= -\frac{1}{4} \sum_{\substack{\mathbf{p}, \mathbf{q} \\ \mathbf{p}', \mathbf{q}'}} \left(\frac{1}{2^{2m}-1} (\text{Tr}[\Omega_{\mathbf{qp}}] \text{Tr}[\Gamma_{\mathbf{pq}}] \text{Tr}[\Omega_{\mathbf{q'p'}}] \text{Tr}[\Gamma_{\mathbf{p'q'}}] + \text{Tr}[\Omega_{\mathbf{qp}} \Omega_{\mathbf{q'p'}}] \text{Tr}[\Gamma_{\mathbf{pq}} \Gamma_{\mathbf{p'q'}}]) \right. \\
&\quad \left. - \frac{1}{2^m(2^{2m}-1)} (\text{Tr}[\Omega_{\mathbf{qp}} \Omega_{\mathbf{q'p'}}] \text{Tr}[\Gamma_{\mathbf{pq}}] \text{Tr}[\Gamma_{\mathbf{p'q'}}] + \text{Tr}[\Omega_{\mathbf{qp}}] \text{Tr}[\Omega_{\mathbf{q'p'}}] \text{Tr}[\Gamma_{\mathbf{pq}} \Gamma_{\mathbf{p'q'}}]) \right) \\
&= -\frac{1}{4(2^{2m}-1)} \sum_{\substack{\mathbf{p}, \mathbf{q} \\ \mathbf{p}', \mathbf{q}'}} \left(\text{Tr}[\Omega_{\mathbf{qp}} \Omega_{\mathbf{q'p'}}] - \frac{1}{2^m} \text{Tr}[\Omega_{\mathbf{qp}}] \text{Tr}[\Omega_{\mathbf{q'p'}}] \right) \text{Tr}[\Gamma_{\mathbf{pq}} \Gamma_{\mathbf{p'q'}}] ,
\end{aligned} \tag{53}$$

where in the third equality we used the fact that the trace of a commutator is zero: $\text{Tr}[\Gamma_{\mathbf{pq}}] = 0$.

If W_B is also a 2-design, then from (53) we need to compute the following expectation value

$$\langle(\partial_\nu C)^2\rangle_{W_B, W_A} = -\frac{1}{4(2^{2m}-1)} \sum_{\substack{\mathbf{p}, \mathbf{q} \\ \mathbf{p}', \mathbf{q}'}} \left(\text{Tr}[\Omega_{\mathbf{qp}} \Omega_{\mathbf{q'p'}}] - \frac{1}{2^m} \text{Tr}[\Omega_{\mathbf{qp}}] \text{Tr}[\Omega_{\mathbf{q'p'}}] \right) \int d\mu(W_B) \text{Tr}[\Gamma_{\mathbf{pq}} \Gamma_{\mathbf{p'q'}}] . \tag{54}$$

Let us first note that

$$\begin{aligned}
\Gamma_{\mathbf{pq}} \Gamma_{\mathbf{p'q'}} &= [\sigma_\nu, W_B \Psi_{\mathbf{pq}} W_B^\dagger] [\sigma_\nu, W_B \Psi_{\mathbf{p'q'}} W_B^\dagger] \\
&= 2 \left(\sigma_\nu W_B \Psi_{\mathbf{pq}} W_B^\dagger \sigma_\nu W_B \Psi_{\mathbf{p'q'}} W_B^\dagger \right) - 2 \left(W_B \sigma_\nu^2 W_B^\dagger \Psi_{\mathbf{pq}} \Psi_{\mathbf{p'q'}} \right) .
\end{aligned} \tag{55}$$

This result can be used along with Lemmas 1 and 2 to compute the integral in (55) as

$$\begin{aligned}
\int d\mu(W_B) \text{Tr}[\Gamma_{\mathbf{pq}} \Gamma_{\mathbf{p'q'}}] &= 2 \int d\mu(W_B) \text{Tr} [\sigma_\nu W_B \Psi_{\mathbf{pq}} W_B^\dagger \sigma_\nu W_B \Psi_{\mathbf{p'q'}} W_B^\dagger] \\
&\quad - 2 \int d\mu(W_B) \text{Tr} [W_B \sigma_\nu^2 W_B^\dagger \Psi_{\mathbf{pq}} \Psi_{\mathbf{p'q'}}] \\
&= -\frac{2^{m+1}}{2^{2m}-1} \text{Tr}[\sigma_\nu^2] \left(\text{Tr}[\Psi_{\mathbf{pq}} \Psi_{\mathbf{p'q'}}] - \frac{1}{2^m} \text{Tr}[\Psi_{\mathbf{pq}}] \text{Tr}[\Psi_{\mathbf{p'q'}}] \right) ,
\end{aligned} \tag{56}$$

where we used the fact that σ_ν is a Pauli operator, and hence its trace is equal to zero.

Then, combining Eqs. (54) and (56), we obtain

$$\langle(\partial_\nu C)^2\rangle_V = \frac{2^{m-1} \text{Tr}[\sigma_\nu^2]}{(2^{2m}-1)^2} \sum_{\substack{\mathbf{pq} \\ \mathbf{p'q'}}} \left\langle \Delta \Omega_{\mathbf{pq}}^{\mathbf{p'q'}} \right\rangle_{V_R} \left\langle \Delta \Psi_{\mathbf{pq}}^{\mathbf{p'q'}} \right\rangle_{V_L} . \tag{57}$$

□

Supplementary Note 5: Variance of the cost function partial derivative for a single layer of the Alternating Layered Ansatz

In this section we explicitly evaluate Eqs. (46)–(48) for the special case when $V(\theta)$ is composed of a single layer of the Alternating Layered Ansatz. This case is a generalization of the warm-up example of the main text, and constitutes the first step towards our main theorems. In particular, we remark that the tools employed here are the same as the ones used to derive our main result.

A. Variance of global cost function partial derivative

Let us first recall that the global cost function is $C_G = 1 - \text{Tr}[OV\rho V^\dagger]$, where $O = \bigotimes_{k=1}^{\xi} \hat{O}_k$, and where $V(\theta)$ is given by a single layer of the Alternating Layered Ansatz, i.e., $V(\theta) = \bigotimes_{k=1}^{\xi} W_{k1}(\theta_k)$. Moreover, we recall that we assume without loss of generality that V_R contains the gates in $\mathcal{L}$ and all the blocks W_{k1} in the last (and in this case only) layer of $V(\theta)$. The latter means that here $V_L = \mathbb{1}$, and $V_R = \mathbb{1}_h \otimes \left(\bigotimes_{k \neq h} W_{k1}(\theta_k) \right)$. In addition, for simplicity, we have defined $W_k := W_{k1}$. Here, ξ is the total number of blocks so that $n = \xi m$, and we assume that the angle θ_ν we want to train is in the h -th block W_h .

1. Expectation value over V_R

First, let us compute Ω_{qp} . From (49), we obtain

$$\Omega_{qp} = \hat{O}_h \prod_{k \neq h}^{\xi} \text{Tr} \left[W_k^\dagger \hat{O}_k W_k |p_k\rangle \langle q_k| \right]. \quad (58)$$

Replacing this result in Eq. (47) and employing Lemma 3 ($\xi - 1$)-times results in

$$\begin{aligned} \langle \Delta \Omega_{pq}^{p'q'} \rangle_{V_R} &= \frac{1}{(2^{2m} - 1)^{\xi-1}} \left(\text{Tr} [\hat{O}_h^2] - \frac{1}{2^m} \text{Tr} [\hat{O}_h]^2 \right) \\ &\times \prod_{k \neq h}^{\xi} \left(\delta_{(p,q)_{S_k}} \delta_{(p',q')_{S_k}} \left(\text{Tr} [\hat{O}_k]^2 - \frac{1}{2^m} \text{Tr} [\hat{O}_k^2] \right) + \delta_{(p',q)_{S_k}} \delta_{(p,q')_{S_k}} \left(\text{Tr} [\hat{O}_k^2] - \frac{1}{2^m} \text{Tr} [\hat{O}_k]^2 \right) \right). \end{aligned}$$

Now, let us consider $\hat{O}_k$ ($k = 1, 2, \dots, \xi$) to be rank-1 projector, i.e. $\text{Tr} [\hat{O}_k] = \text{Tr} [\hat{O}_k^2] = \text{rank} [\hat{O}_k] = 1$. Therefore, we can obtain

$$\langle \Delta \Omega_{pq}^{p'q'} \rangle_{V_R} = \frac{1}{(2^{2m} - 1)^{\xi-1}} \left(1 - \frac{1}{2^m} \right)^{\xi} \prod_{k \neq h}^{\xi} \left(\delta_{(p,q)_{S_k}} \delta_{(p',q')_{S_k}} + \delta_{(p',q)_{S_k}} \delta_{(p,q')_{S_k}} \right). \quad (59)$$

2. Expectation value over V_L

Next, let us consider Ψ_{pq} in (50). Here, we can set $V_L = \mathbb{1}$, which leads to $\Psi_{pq} = \text{Tr}_h[(|p\rangle \langle q| \otimes \mathbb{1}_h) \rho]$. Note that any quantum state ρ can be always written as

$$\rho = \sum_{\lambda} p_{\lambda} |\psi_{\lambda}\rangle \langle \psi_{\lambda}|, \quad \text{with} \quad |\psi_{\lambda}\rangle = \sum_{\alpha^{\lambda}} c_{\alpha^{\lambda}} |\alpha_1^{\lambda}\rangle \otimes \dots \otimes |\alpha_h^{\lambda}\rangle \otimes \dots \otimes |\alpha_{\xi}^{\lambda}\rangle, \quad (60)$$

where $\alpha := \alpha_1 \dots \alpha_{\xi}$, and where α_i are bitstrings of length m . Hence we find

$$\Psi_{pq} = \sum_{\lambda} p_{\lambda} \sum_{\alpha^{\lambda}, \alpha'^{\lambda}} c_{\alpha^{\lambda}} c_{\alpha'^{\lambda}}^* \left(\prod_{k \neq h}^{\xi} \delta_{(q, \alpha^{\lambda})_k} \delta_{(p, \alpha'^{\lambda})_{S_k}} \right) |\alpha_h^{\lambda}\rangle \langle \alpha'^{\lambda}_h|. \quad (61)$$

Then, since $V_L = \mathbb{1}$, we have $\langle \Delta \Psi_{pq}^{p'q'} \rangle_{V_L} = \Delta \Psi_{pq}^{p'q'}$, and we can use (61) to get

$$\begin{aligned} \Delta \Psi_{pq}^{p'q'} &= \sum_{\lambda, \lambda'} p_\lambda p_{\lambda'} \sum_{\substack{\alpha^\lambda, \alpha'^\lambda \\ \beta^{\lambda'}, \beta'^{\lambda'}}} c_{\alpha^\lambda} c_{\alpha'^\lambda}^* c_{\beta^{\lambda'}} c_{\beta'^{\lambda'}}^* \left(\prod_{k \neq h}^{\xi} \delta_{(q, \alpha^\lambda)_{S_k}} \delta_{(p, \alpha'^\lambda)_{S_k}} \delta_{(q', \beta^{\lambda'})_{S_k}} \delta_{(p', \beta'^{\lambda'})_{S_k}} \right) \\ &\quad \times \left(\delta_{(\alpha'^\lambda, \beta^{\lambda'})_{S_h}} \delta_{(\alpha^\lambda, \beta'^{\lambda'})_{S_h}} - \frac{1}{2^m} \delta_{(\alpha^\lambda, \alpha'^\lambda)_{S_h}} \delta_{(\beta^{\lambda'}, \beta'^{\lambda'})_{S_h}} \right). \end{aligned} \quad (62)$$

Finally, from (59), (62), and the fact that $\text{Tr}[\sigma_\nu^2] = 2^m$, we obtain

$$\begin{aligned} \text{Var}[\partial_\nu C_G] &= \frac{2^{2m-1}}{(2^{2m}-1)^2(2^{2m}-1)^{\xi-1}} \left(1 - \frac{1}{2^m} \right)^\xi \sum_{\lambda, \lambda'} p_\lambda p_{\lambda'} \sum_{\substack{\alpha^\lambda, \alpha'^\lambda \\ \beta^{\lambda'}, \beta'^{\lambda'}}} c_{\alpha^\lambda} c_{\alpha'^\lambda}^* c_{\beta^{\lambda'}} c_{\beta'^{\lambda'}}^* \\ &\quad \times \left(\delta_{(\alpha'^\lambda, \beta^{\lambda'})_{S_h}} \delta_{(\alpha^\lambda, \beta'^{\lambda'})_{S_h}} - \frac{1}{2^m} \delta_{(\alpha^\lambda, \alpha'^\lambda)_{S_h}} \delta_{(\beta^{\lambda'}, \beta'^{\lambda'})_{S_h}} \right) \prod_{k \neq h} \left(\delta_{(\alpha^\lambda, \alpha'^\lambda)_{S_k}} \delta_{(\beta^{\lambda'}, \beta'^{\lambda'})_{S_k}} + \delta_{(\alpha^\lambda, \beta'^{\lambda'})_{S_k}} \delta_{(\alpha'^\lambda, \beta^{\lambda'})_{S_k}} \right), \end{aligned}$$

where we used the fact that

$$\begin{aligned} &\sum_{\substack{pq \\ p'q'}} \left(\prod_{k \neq h}^{\xi} \left(\delta_{(p, q)_{S_k}} \delta_{(p', q')_{S_k}} + \delta_{(p', q)_{S_k}} \delta_{(p, q')_{S_k}} \right) \delta_{(q, \alpha^\lambda)_{S_k}} \delta_{(p, \alpha'^\lambda)_{S_k}} \delta_{(q', \beta^{\lambda'})_{S_k}} \delta_{(p', \beta'^{\lambda'})_{S_k}} \right) \\ &= \prod_{k \neq h} \left(\delta_{(\alpha^\lambda, \alpha'^\lambda)_{S_k}} \delta_{(\beta^{\lambda'}, \beta'^{\lambda'})_{S_k}} + \delta_{(\alpha^\lambda, \beta'^{\lambda'})_{S_k}} \delta_{(\alpha'^\lambda, \beta^{\lambda'})_{S_k}} \right). \end{aligned}$$

Let us define

$$\begin{aligned} J &:= \sum_{\lambda, \lambda'} p_\lambda p_{\lambda'} \sum_{\substack{\alpha^\lambda, \alpha'^\lambda \\ \beta^{\lambda'}, \beta'^{\lambda'}}} c_{\alpha^\lambda} c_{\alpha'^\lambda}^* c_{\beta^{\lambda'}} c_{\beta'^{\lambda'}}^* \\ &\quad \times \left(\delta_{(\alpha'^\lambda, \beta^{\lambda'})_{S_h}} \delta_{(\alpha^\lambda, \beta'^{\lambda'})_{S_h}} - \frac{1}{2^m} \delta_{(\alpha^\lambda, \alpha'^\lambda)_{S_h}} \delta_{(\beta^{\lambda'}, \beta'^{\lambda'})_{S_h}} \right) \prod_{k \neq h} \left(\delta_{(\alpha^\lambda, \alpha'^\lambda)_{S_k}} \delta_{(\beta^{\lambda'}, \beta'^{\lambda'})_{S_k}} + \delta_{(\alpha^\lambda, \beta'^{\lambda'})_{S_k}} \delta_{(\alpha'^\lambda, \beta^{\lambda'})_{S_k}} \right), \end{aligned}$$

which, has the form of $J = 1 - \frac{1}{2^m} + \sum_{l=1}^{2^{\xi-1}-1} (\mathcal{A}_L - \frac{1}{2^m} \mathcal{B}_L)$, where $\mathcal{A}_L, \mathcal{B}_L$ are the purities of reduced states of ρ . The latter can be understood in the following way. Let us define the set $\mathcal{S} = \{S_1, S_2, \dots, S_\xi\}$, whose elements represents each block. Then, suppose that we want to consider the partial trace of ρ over several subsystems $\mathcal{H}_{\overline{K}}$, where $\overline{K} \subset \mathcal{S}$. The reduced state lives in the composite Hilbert space $\mathcal{H}_K$, where $K = \mathcal{S} \setminus \overline{K}$, and we can write $\alpha := \alpha_K \cdot \alpha_{\overline{K}}$. Let us define the reduced state as $\rho_K := \text{Tr}_{\overline{K}}[\rho]$, and its purity can be explicitly written as

$$\text{Tr}[\rho_K^2] = \sum_{\lambda, \lambda'} p_\lambda p_{\lambda'} \sum_{\alpha^\lambda, \beta^{\lambda'}} c_{\alpha^\lambda} c_{\alpha_{\overline{K}}^\lambda}^* c_{\beta^{\lambda'}} c_{\beta_{\overline{K}}^{\lambda'}}^* c_{\alpha_K^\lambda} c_{\alpha_{\overline{K}}^\lambda}^* c_{\beta_K^{\lambda'}} c_{\beta_{\overline{K}}^{\lambda'}}^*,$$

which appears in the expression of J . Since $\text{Tr}[\rho_K^2] \leq 1$, we have $J < 2^{\xi-1}$; therefor, we can write

$$\text{Var}[\partial_\nu C_G] < \frac{2^{2m-2}}{(2^{2m}-1)2^{2m\xi-\xi}(1+2^{-m})^\xi} < \frac{1}{2^{(2-\frac{1}{m})n}} \quad (\forall m \in \mathbb{N}), \quad (63)$$

where in the last inequality we used the fact that $n = m\xi$. Equation (63) shows that for the single layer of the alternating layered ansatz the global cost functions presents a barren plateau.

B. Variance of the local cost function partial derivative

Here we consider the case when $V(\theta)$ is given by a single layer of the Alternating Layered Ansatz, and when the local cost function is

$$C_L = 1 - \frac{1}{n} \sum_{j=1}^n \text{Tr} \left[\left(|0\rangle\langle 0|_j \otimes \mathbb{1}_{\overline{j}} \right) V \rho V^\dagger \right].$$

with $O_L = \frac{1}{n} \sum_{j=1}^n |0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}}$, and where j denotes the j -th qubit. Moreover, let us assume that we are training a parameter θ_ν in h -th block. This case is simpler than the one previously considered for the global cost function as the gradient is non vanishing only when we measure a qubit j in S_h . We can then redefine $O_L^h : \mathcal{H}_h \rightarrow \mathcal{H}_h$ as $\hat{O}_L^h = \frac{1}{n} \sum_{j=1}^m |0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}}$, where each j is such that $j \in S_h$. Then, from (57) we have

$$\text{Tr}[\hat{O}_L^h] = \frac{m2^{m-1}}{n}, \quad \text{Tr}\left[\left(\hat{O}_L^h\right)^2\right] = \frac{m(m+1)2^{m-2}}{n^2}. \quad (64)$$

From (49), Ω_{qp} can be written as $\Omega_{qp} = \delta_{pq} W_h O_L^h W_h^\dagger$, which leads to

$$\Delta\Omega_{pq}^{p'q'} = \frac{m2^{m-2}}{n^2} \delta_{pq} \delta_{p'q'}. \quad (65)$$

Next, from (50) we have $\Psi_{pq} = \text{Tr}_{\bar{h}}[|q\rangle\langle p| \otimes \mathbb{1}_h] \rho$, and it is straightforward to see that

$$\sum_{pq} \delta_{pq} \Psi_{pq} = \sum_{pq} \delta_{pq} \text{Tr}_{\bar{h}}[|q\rangle\langle p| \otimes \mathbb{1}_h] \rho = \rho_h, \quad (66)$$

where we defined $\rho_h := \text{Tr}_{\bar{h}}[\rho]$. Then, from (57), and by using $\left\langle \Delta\Psi_{pq}^{p'q'} \right\rangle_{V_L} = \Delta\Psi_{pq}^{p'q'}$, and $\text{Tr}[\sigma_\nu^2] = 2^m$, we can write

$$\text{Var}[\partial_\nu C_L] = \frac{m2^{3(m-1)}}{n^2(2^{2m}-1)^2} \left(\text{Tr}[\rho_h^2] - \frac{1}{2^m} \right) = \frac{m \cdot 2^{3(m-1)}}{n^2(2^{2m}-1)^2} D_{HS} \left(\rho_h, \frac{\mathbb{1}}{2^m} \right),$$

where $D_{HS}(\rho_h, \mathbb{1}/2^m) = \text{Tr}[(\rho_h - \mathbb{1}/2^m)^2]$ is the Hilbert-Schmidt distance between ρ_h and $\mathbb{1}/2^m$. If $D_{HS}(\rho_h, \mathbb{1}/2^m) \in \Omega(1/\text{poly}(n))$, we have that the variance of the cost function partial derivative is polynomially vanishing with n as

$$\text{Var}[\partial_\nu C_L] \in \Omega\left(\frac{1}{\text{poly}(n)}\right), \quad (67)$$

and hence in this case C_L presents no barren plateau.

Supplementary Note 6: Proof of Theorem 2

First, let us recall that we are considering m -local cost functions where each operator O_i acts nontrivially on m qubits $O_i = \mathbb{1}_{\bar{m}} \otimes \hat{O}_i$ (here $\mathbb{1}_{\bar{m}}$ indicates the identity on all but m qubits), and where $\hat{O}_i$ can be expressed as $\hat{O}_i = \hat{O}_i^{\mu_i} \otimes \hat{O}_i^{\mu'_i}$. Hence, we have

$$O = c_0 \mathbb{1} + \sum_{i=1}^N c_i \hat{O}_i^{\mu_i} \otimes \hat{O}_i^{\mu'_i}, \quad (68)$$

where $\hat{O}_i^{\mu_i}$ are operators acting on $m/2$ qubits which can be written as tensor product of Pauli operators. Here we recall that we have defined S_k as the m -qubit subsystem on which W_{kL} acts and let $\mathcal{S} = \{S_k\}$ be the set of all such subsystems. As detailed in the main text the summation in Eq. (68) includes two possible cases: First, when $\hat{O}_i^{\mu_i}$ ($\hat{O}_i^{\mu'_i}$) acts on the first (last) $m/2$ qubits of a given S_k , and second, when $\hat{O}_i^{\mu_i}$ ($\hat{O}_i^{\mu'_i}$) acts on the last (first) $m/2$ qubits of a given S_k (S_{k+1}). This type of cost function includes any ultralocal (i.e., where the O_i are one-body) cost function, and also VQE Hamiltonians with up to $m/2$ neighbor interactions.

Here we provide a proof of Theorem 2 in the main text, which we now reiterate for convenience.

Theorem 2. Consider a trainable parameter θ^ν in a block W of the ansatz in Fig. 3 of the main text. Let $\text{Var}[\partial_\nu C]$ be the variance of the partial derivative of an m -local cost function C (with O given by (68)) with respect to θ^ν . If W_A , W_B of (39), and each block in $V(\theta)$ form a local 2-design, then $\text{Var}[\partial_\nu C]$ is lower bounded by

$$G_n(L, l) \leq \text{Var}[\partial_\nu C], \quad (69)$$

with

$$G_n(L, l) = \frac{2^{m(l+1)-1}}{(2^{2m}-1)^2(2^m+1)^{L+l}} \sum_{i \in \mathcal{L}} \sum_{\substack{(k,k') \in \mathcal{L}_B \\ k' \geq k}} c_i^2 \epsilon(\rho_{k,k'}) \epsilon(\hat{O}_i), \quad (70)$$

where $i_{\mathcal{L}}$ is the set of i indices whose associated operators $\hat{O}_i$ act on qubits in the forward light-cone $\mathcal{L}$ of W , and $k_{\mathcal{L}_B}$ is the set of k indices whose associated subsystems S_k are in the backward light-cone $\mathcal{L}_B$ of W . Here we defined the function $\epsilon(M) = D_{\text{HS}}(M, \text{Tr}(M)\mathbb{1}/d_M)$ where D_{HS} is the Hilbert-Schmidt distance and d_M is the dimension of the matrix M . In addition, $\rho_{k,k'}$ is partial trace of the input state ρ down to the subsystems $S_k S_{k+1} \dots S_{k'}$.

Proof. Let us first consider the case when the operators O_i are of the form (68) and act non trivially in a given subsystem of $\mathcal{S}$. We can expand

$$\text{Var}[\partial_\nu C] = \frac{2^{m-1} \text{Tr}[\sigma_\nu^2]}{(2^{2m} - 1)^2} \sum_{i,j} \sum_{\substack{\mathbf{p}\mathbf{q} \\ \mathbf{p}'\mathbf{q}'}} c_i c_j \left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]}{2^m} \right\rangle_{V_R} \left\langle \Delta \Psi_{\mathbf{p}\mathbf{q}}^{\mathbf{p}'\mathbf{q}'} \right\rangle_{V_L}, \quad (71)$$

where we defined

$$\Omega_{\mathbf{q}\mathbf{p}}^i = \text{Tr}_{\bar{w}} \left[(|\mathbf{p}\rangle\langle\mathbf{q}| \otimes \mathbb{1}_w) V_R^\dagger O_i V_R \right]. \quad (72)$$

1. Expectation value over V_R

We recall here that we assume without loss of generality that V_R contains the gates in $\mathcal{L}$ and all the blocks W_{kL} in the last layer of $V(\theta)$. Hence, we can express

$$V_R = V_{\bar{\mathcal{L}}} \otimes V_{\mathcal{L}}, \quad (73)$$

where $V_{\mathcal{L}}$ contains all the blocks in the forward light-cone $\mathcal{L}$ of W , and where $V_{\bar{\mathcal{L}}}$ consists of all the remaining blocks in the last layer of the ansatz. When analyzing the operators $\Omega_{\mathbf{p}\mathbf{q}}^i$ we have to consider two cases: 1) when O_i only acts non-trivially on qubits in $S_{\bar{\mathcal{L}}}$, and 2) when $\hat{O}_i$ acts non-trivially on qubits in $S_{\mathcal{L}}$. If O_i only acts non-trivially on qubits in $S_{\bar{\mathcal{L}}}$, it is straightforward to show from (72) that $\Omega_{\mathbf{p}\mathbf{q}}^i \propto \mathbb{1}_w$. Hence, in this case, we find

$$\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]}{2^m} = 0. \quad (74)$$

Let us then define $i_{\mathcal{L}}$ as the set of i indices whose associated operators O_i act on qubits in the forward light-cone $\mathcal{L}$ of W . In what follows we assume that the indexes $i, j \in i_{\mathcal{L}}$, i.e., we assume that O_i and O_j act non-trivially on qubit in $S_{\mathcal{L}}$. The latter leads to

$$\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]}{2^m} \right\rangle_{V_R} = \delta_{(\mathbf{p},\mathbf{q})_{S_{\bar{\mathcal{L}}}}} \delta_{(\mathbf{p}',\mathbf{q}')_{S_{\bar{\mathcal{L}}}}} \left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]}{2^m} \right\rangle_{V_{\mathcal{L}}}. \quad (75)$$

Here the delta functions arise from (72) by noting that $V_R^\dagger O_i V_R = V_{\bar{\mathcal{L}}}^\dagger O_i V_{\bar{\mathcal{L}}}$.

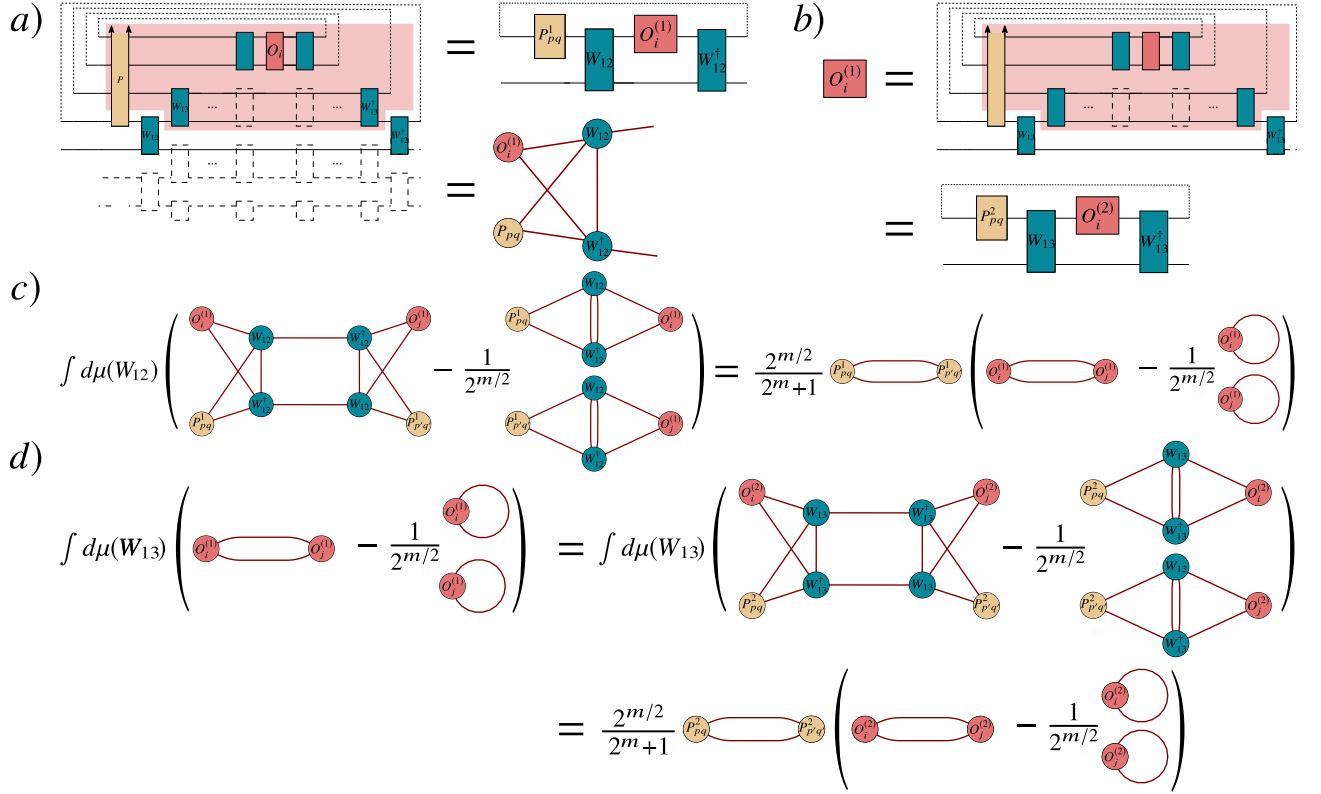
In order to explicitly evaluate the expectation value in (75) we use the fact that each block in the layered ansatz is a 2-design, and hence one can algorithmically integrate over each block using the Weingarten calculus. Specifically, as discussed in the main text we employ the tensor network representation of $\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j]$ and $\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]$, and we use the *Random Tensor Network Integrator* (RTNI) package of Ref. [5], which allows for the computation of averages of tensor networks containing multiple Haar-distributed random unitary matrices and deterministic symbolic tensors.

Using this procedure, $\langle \dots \rangle_{V_{\mathcal{L}}}$ can be computed by integrating over each block inside of $V_{\mathcal{L}}$ over the unitary group with respect to the Haar measure. After each integration the result of the average is a sum of up to four new tensor according to (3), and Lemmas 2 and 3. After all the blocks in $V_{\mathcal{L}}$ have been integrated, the final result can be expressed as

$$\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]}{2^m} \right\rangle_{V_{\mathcal{L}}} = \sum_{\tau} t_{\tau}^{ij} \delta_{(\mathbf{p},\mathbf{q})_{S_{\tau}}} \delta_{(\mathbf{p}',\mathbf{q}')_{S_{\bar{\tau}}}} \delta_{(\mathbf{p},\mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}',\mathbf{q})_{S_{\bar{\tau}}}} \Delta O_{\tau}^{ij}, \quad (76)$$

where $t_{\tau} \in \mathbb{R}$, $S_{\tau} \cup S_{\bar{\tau}} = S_{\mathcal{L}} \cap S_{\bar{w}}$ (with $S_{\tau} \neq \emptyset$), and where we have defined

$$\Delta O_{\tau}^{ij} = \text{Tr}_{x_{\tau} y_{\tau}} [\text{Tr}_{z_{\tau}} [O_i] \text{Tr}_{z_{\bar{\tau}}} [O_j]] - \frac{\text{Tr}_{x_{\tau}} [\text{Tr}_{y_{\tau} z_{\tau}} [O_i] \text{Tr}_{y_{\tau} z_{\bar{\tau}}} [O_j]]}{2^m}. \quad (77)$$



Supplementary Figure 1. Schematic representation of the method employed to derive Eq. (76). Here we consider the case when O_i and O_j act nontrivially only on the topmost m qubits in $V_{\mathcal{L}}$. (a) Tensor-network representation of the term Ω_{qp}^i . All but $L - l$ blocks in $V_{\mathcal{L}}$ simplify to identity when computing $V_{\mathcal{L}}^\dagger O_i V_{\mathcal{L}}$. One can always define operators $O_i^{(1)}$, and $O_j^{(1)}$ such that Eq. (78) holds. Here W_{12} is the topmost block in the first layer of $V_{\mathcal{L}}$, and we define P_{pq}^1 as a projector operator acting on $m/2$ qubits. (b) Tensor-network representation of the operator $O_i^{(1)}$. Note that from $O_i^{(1)}$ we can also define an operator $O_i^{(2)}$ which will be of the same form as $O_i^{(1)}$. (c) By means of the Weingarten calculus we can integrate over W_{12} to compute the expectation value of Eq. (78). The result is a single tensor. This result can be obtained by employing Eq. (23) of Lemma 6. (d) In order to integrate the next block we use the operators $O_i^{(2)}$ to express the tensor networks in a form that coincides with panel (c). We can then repeat this procedure $L - l$ times in order to integrate over all blocks and compute the expectation value in (75). The final result is of the form of Eq. (79).

Here we use the notation Tr_{x_τ} to indicate the trace over the Hilbert space associated with subsystem S_{x_τ} . We also define $\mathcal{S}^{\mathcal{L}} = \{S_k : S_k \subset \mathcal{S}^{\mathcal{L}}\}$ as the set of subspaces S_k which belong to the light-cone $\mathcal{L}$, so that we have $S_{y_\tau} \in \mathcal{S}^{\mathcal{L}}$, $S_{x_\tau} \cup S_{y_\tau} \cup S_{z_\tau} = \mathcal{S}^{\mathcal{L}}$, and $S_{x_\tau}, S_{z_\tau} \in \mathcal{P}(\mathcal{S}^{\mathcal{L}})$, with $\mathcal{P}(\mathcal{S}^{\mathcal{L}}) = \{\emptyset, S_1, \dots, S_\xi, S_1 \cup S_2, \dots\}$ the power set of $\mathcal{S}^{\mathcal{L}}$. If one chooses $S_{z_\tau} = \emptyset$ (or $S_{x_\tau} = \emptyset$), with associated Hilbert space $H_0 = \{\mathbf{0}\}$, we define $\text{Tr}_\emptyset [O] := O$.

Equation (76) can be derived by iteratively applying Lemma 6 each time a block in $V_{\mathcal{L}}$ is integrated in (75), and noting that at each integration the resulting tensor can always be written in a way such that Lemma 6 can be applied again. As an example, let us consider the case when O_i and O_j act nontrivially only on the topmost m qubits in $V_{\mathcal{L}}$. As shown in Sup. Fig. 1(a), all but $L - l$ blocks in $V_{\mathcal{L}}$ simplify to identity when computing $V_{\mathcal{L}}^\dagger O_i V_{\mathcal{L}}$. Then, let us denote by W_{12} the topmost gate in the first layer in $V_{\mathcal{L}}$, and let $\tilde{S}_1, \tilde{S}_2$ be the sets of $m/2$ adjacent qubits (with associated Hilbert spaces $\tilde{\mathcal{H}}_1, \tilde{\mathcal{H}}_2$) such that W_{12} acts on $\tilde{\mathcal{H}}_1 \otimes \tilde{\mathcal{H}}_2$, and such that $\tilde{S}_2 \subset S_w$. One can always write

$$\left\langle \text{Tr}[\Omega_{qp}^i \Omega_{q'p'}^j] - \frac{\text{Tr}[\Omega_{qp}^i] \text{Tr}[\Omega_{q'p'}^j]}{2^m} \right\rangle_{V_{\mathcal{L}}} \propto \left\langle \text{Tr}_{\tilde{\mathcal{H}}_2} \left[\text{Tr}_{\tilde{\mathcal{H}}_1} [\Omega_1] \text{Tr}_{\tilde{\mathcal{H}}_1} [\Omega'_1] \right] - \frac{\text{Tr}[\Omega_1] \text{Tr}[\Omega'_1]}{2^m} \right\rangle_{V_{\mathcal{L}}}, \quad (78)$$

with $\Omega_1 = W_{12} O_i^{(1)} W_{12}^\dagger (|\mathbf{p}\rangle \langle \mathbf{q}|_{\tilde{S}_1} \otimes \mathbb{1}_{\tilde{S}_2})$, $\Omega'_1 = W_{12} O_j^{(1)} W_{12}^\dagger (|\mathbf{p}'\rangle \langle \mathbf{q}'|_{\tilde{S}_1} \otimes \mathbb{1}_{\tilde{S}_2})$, and where $O_i^{(1)}$, and $O_j^{(1)}$ are defined according to Sup. Fig. 1(a). The proportionality factor in Eq. (78) is given by delta functions over $\mathbf{p}$, and $\mathbf{q}$. Moreover, the right-hand side of (78) is exactly of the form (16), and hence by applying Lemma 6 (or more specifically, Eq. (23)) one finds the result of Sup. Fig. 1(c). As schematically depicted in Sup. Fig. 1(b), one can then define operators $O_i^{(2)}$

and repeat this calculation $L - l$ times as in Sup. Fig. 1(d). One finally finds

$$\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]}{2^m} \right\rangle_{V_{\mathcal{L}}} \propto \frac{2^{m(L-l)/2}}{(2^m + 1)^{L-l}} \left(\text{Tr}[O_i O_j] - \frac{\text{Tr}[O_i] \text{Tr}[O_j]}{2^m} \right), \quad (79)$$

where once again the proportionally factor is given by delta functions over $\mathbf{p}$ and $\mathbf{q}$. For more general operators O_i and O_j , each time a block is integrated, one can always rewrite the resulting non-trivial terms in the form of either Eq. (16), or Eq. (20). Remarkably, this means that the final result can always be expressed in terms of contractions of the form (77), and hence Eq. (76) holds.

As previously mentioned, Eq. (76) is valid for arbitrary operators O_i and O_j . However, from Lemma 7 we know that if O_i and O_j have no overlapping support, then $\Delta O_{\tau}^{ij} = 0$, for all τ . Hence we only have to consider the cases when $i = j$. Moreover, if $\hat{O}_i$ act non trivially in a given subsystem of $\mathcal{S}$ the summation in Eq. (76) simplifies and we find

$$\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]}{2^m} \right\rangle_{V_{\mathcal{L}}} = \epsilon(\hat{O}_i) \sum_{\tau} \hat{t}_{\tau}^{ii} \delta_{(\mathbf{p}, \mathbf{q})_{S_{\overline{\tau}}}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_{\overline{\tau}}}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_{\tau}}}, \quad (80)$$

where we now denote the coefficients as $\hat{t}_{\tau}^{ii}$ since we now have $\hat{t}_{\tau}^{ii} \geq 0$, and where

$$\epsilon(\hat{O}_i) = \text{Tr}[\hat{O}_i^2] - \frac{\text{Tr}[\hat{O}_i]^2}{2^m} = D_{HS} \left(\hat{O}_i, \text{Tr}[\hat{O}_i] \frac{\mathbb{1}}{2^m} \right). \quad (81)$$

Moreover, we also find that the following inequality holds $\forall i$

$$\sum_{\tau} \hat{t}_{\tau}^{ii} \geq \frac{2^{m(L-l)/2}}{(2^m + 1)^{L-l}}, \quad (82)$$

where we recall that L is the number of layers in the ansatz, and that the block W is in the l th-layer of $V(\theta)$.

Combining Eqs. (71) and (80), we find

$$\text{Var}[\partial_{\nu} C] = \frac{2^{m-1} \text{Tr}[\sigma_{\nu}^2]}{(2^{2m} - 1)^2} \sum_{\substack{\mathbf{p}\mathbf{q} \\ \mathbf{p}'\mathbf{q}'}} \sum_{i \in i_{\mathcal{L}}} c_i^2 \epsilon(\hat{O}_i) \sum_{\tau} \hat{t}_{\tau}^{ii} \delta_{(\mathbf{p}, \mathbf{q})_{S_{\overline{\mathcal{L}} \cup S_{\overline{\tau}}}}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_{\overline{\mathcal{L}} \cup S_{\overline{\tau}}}}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_{\tau}}} \left\langle \Delta \Psi_{\mathbf{p}\mathbf{q}}^{\mathbf{p}'\mathbf{q}'} \right\rangle_{V_{\mathcal{L}}}. \quad (83)$$

Before proceeding to compute the expectation value over $V_{\mathcal{L}}$, let us analyze the lower bound in (82). For simplicity of notation, let us define $T_{\tau} := \sum_{\tau} \hat{t}_{\tau}^{ii}$. In what follows we compare T_{τ} for two relevant (and extremum) cases: 1) when O_i acts nontrivially on the topmost m qubits in $V_{\mathcal{L}}$ (see panel (a) of Fig. (2)), and 2) when O_i acts nontrivially on S_w (see panel (b) of Fig. (2)). As discussed below, T_{τ} is larger for Case 2, that for Case 1.

As schematically depicted in Sup. Fig. (2)(a), for Case 1 one needs to integrate over $(L - l)$ blocks, and we know from Eq. (79) (and Sup. Fig. 1) that

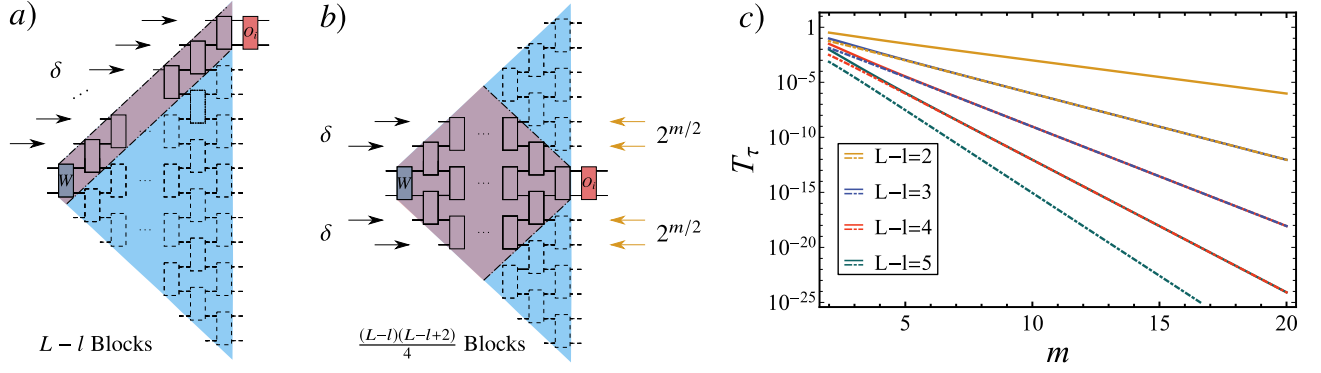
$$T_{\tau} = \frac{2^{m(L-l)/2}}{(2^m + 1)^{L-l}}. \quad (84)$$

On the other hand, for Case 2 one needs to integrate over $(L - l)(L - l + 2)/4$ blocks. While a closed formula cannot be derived for this case, we can explicitly integrate over each block to obtain:

$$T_{\tau} = \begin{cases} \frac{2^{3m/2}(1+2^{m/2})^2}{(1+2^m)^3} & \text{if } L - l = 2 \\ \frac{2^{2m}(1+2^{m/2})^2(1+2^{m/2}+2^m)}{(1+d^2)^5} & \text{if } L - l = 3 \\ \frac{2^{5m/2}(1+2^{m/2}+2^m)(1+2^{m/2}(3+2^{m/2}(8+2^{m/2}(3+2^{m/2}))))}{(1+2^m)^7} & \text{if } L - l = 4 \\ \frac{2^{3m}(1+2^{m/2}(2+2^{m/2}(4+2^{m/2}(2+2^{m/2}(3+2^{m/2}(8+2^{m/2}(3+2^{m/2}))))))}{(1+2^m)^9} & \text{if } L - l = 5 \end{cases}. \quad (85)$$

For all cases we can see from Sup. Fig. (2)(c) that the values of T_{τ} for Case 2 are larger than the ones obtained for Case 1.

The latter can be understood from two key facts. First, that while integrating over a block leads to coefficients smaller than one according to (3), the more blocks one integrates over, the more contributions we get for $\epsilon(\hat{O}_i)$.



Supplementary Figure 2. a) Schematic representation of Case 1. The operator O_i acts nontrivially on the m topmost qubits in V_L . Note that this scenario corresponds to the one analyzed in Sup. Fig. 1. Here one needs to integrate over $(L-l)$ blocks. The arrows indicate the tensor indexes that, when contracted after integrating over each block, will lead to delta functions in $\mathbf{p}$ and $\mathbf{q}$. b) Schematic representation of Case 2. The operator O_i acts non trivially on S_w . Here one needs to integrate over $(L-l)(L-l+2)/4$ blocks. For the case of L odd, the operator O_i acts nontrivially on the top $m/2$ qubits of S_w and the $m/2$ qubits above those. The arrows indicate the tensor indexes that, when contracted after integrating over each block, will lead to either delta functions in $\mathbf{p}$ and $\mathbf{q}$, or to traces over the identity on $m/2$ qubits. c) Solid lines represent the value of T_τ versus m for Case 2 in panel b) (i.e., from Eq. (85)). Dashed lines represent the value of T_τ versus m for Case 1 in panel a) (i.e., from Eq. (84)). The y axis is shown in a logarithmic scale. For all values of $(L-l)$ considered the values of T_τ Case 2 are larger than those Case 1. For a given value of $(L-l)$ the difference between values of T_τ in (84) and in (85) scales as $1/2^{m(L-l-1)/2}$. The previous explains why as m increases we the dashed line for $(L-l)$ converges to the solid line with $(L-l+1)$.

Second, from the fact that in Case 2 the integration over some blocks will lead to multiplicative factors $2^{m/2}$ which will compensate the factors from (3). These large factors $2^{m/2}$ essentially arise when integrating blocks in V_L which act on qubits which one does not measure, and which hence lead to traces over the identity over $m/2$ qubits. Note that these factors do not arise in Case I. In Sup. Fig. (2)(a) we have schematically indicated by arrows the indexes that will lead to delta functions in $\mathbf{p}$ and $\mathbf{q}$ after each block is integrated. Each time we integrate a block containing these indexes one obtains tensor contractions of $|\mathbf{p}\rangle\langle\mathbf{q}|$. On the other hand, for Case 2, one will not only get these tensor contractions in $\mathbf{p}$ and $\mathbf{q}$, but as indicated in Sup. Fig. (2)(b) we have also indexes that will lead to identity-tensor contractions and hence to factors $2^{m/2}$.

In fact, for a given L the difference between the T_τ in (84) and in (85) scales as $1/2^{m(L-l-1)/2}$, which means that as $(L-l)$ increases the inequality in (82) becomes tighter. Finally, we remark that the previous explains why in Fig. (2)(c) the values of T_τ from Case 1 and fixed $(L-l)$ converge to the values of Case 2 and $(L-l+1)$.

2. Expectation value over V_L

Let us now consider the term

$$\left\langle \Delta \Psi_{\mathbf{p}\mathbf{q}}^{\mathbf{p}'\mathbf{q}'} \right\rangle_{V_L} = \left\langle \text{Tr}[\Psi_{\mathbf{p}\mathbf{q}} \Psi_{\mathbf{p}'\mathbf{q}'}] - \frac{\text{Tr}[\Psi_{\mathbf{p}\mathbf{q}}] \text{Tr}[\Psi_{\mathbf{p}'\mathbf{q}'}]}{2^m} \right\rangle_{V_L}. \quad (86)$$

Since the bitstring $\mathbf{p}$ ($\mathbf{q}$) can be expressed as a bit-wise concatenation of the form $\mathbf{p} = (\mathbf{p})_{S_{\bar{L}} \cup S_{\bar{\tau}}} \cdot (\mathbf{p})_{S_{\tau}}$ (and similarly for $\mathbf{q}$), we can first evaluate

$$\sum_{(\mathbf{p}\mathbf{q})_{S_{\bar{L}} \cup S_{\bar{\tau}}}} \delta_{(\mathbf{p},\mathbf{q})_{S_{\bar{L}} \cup S_{\bar{\tau}}}} \Psi_{\mathbf{p}\mathbf{q}} = \text{Tr}_{\tau} [(|\mathbf{p}\rangle\langle\mathbf{q}|_{S_{\tau}} \otimes \mathbb{1}_w) \tilde{\rho}_{Wk}], \quad (87)$$

where Tr_{τ} is the partial trace over the Hilbert space $\mathcal{H}_{\tau}$ of the qubits in S_{τ} , and where we defined

$$\tilde{\rho}_{Wk} = \text{Tr}_{\bar{L} \cup \bar{\tau}} [V_L \rho V_L^{\dagger}], \quad (88)$$

as the reduced state of $V_L \rho V_L^{\dagger}$ on $\mathcal{H}_w \otimes \mathcal{H}_{\tau}$. Then, from the term $\delta_{(\mathbf{p},\mathbf{q})_{S_{\tau}}} \delta_{(\mathbf{p}',\mathbf{q}')_{S_{\tau}}}$ in (83), we get

$$\sum_{(\mathbf{p}\mathbf{q})_{S_{\tau}}} \delta_{(\mathbf{p},\mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}',\mathbf{q})_{S_{\tau}}} \left(\text{Tr}[\Psi_{\mathbf{p}\mathbf{q}} \Psi_{\mathbf{q}\mathbf{p}}] - \frac{\text{Tr}[\Psi_{\mathbf{p}\mathbf{q}}] \text{Tr}[\Psi_{\mathbf{q}\mathbf{p}}]}{2^m} \right) = D_{HS} \left(\tilde{\rho}_{w\tau}, \tilde{\rho}_{\tau} \otimes \frac{\mathbb{1}}{2^m} \right), \quad (89)$$

where $D_{HS}(\rho, \sigma) = \text{Tr}[(\rho - \sigma)^2]$ is the Hilbert-Schmidt distance, and where we defined

$$\tilde{\rho}_\tau = \text{Tr}_w(\tilde{\rho}_{w\tau}), \quad \text{and} \quad \tilde{\rho}_w = \text{Tr}_\tau(\tilde{\rho}_{w\tau}) = \text{Tr}_{\bar{w}}[V_L \rho V_L^\dagger], \quad (90)$$

as the reduces states of $V_L \rho V_L^\dagger$ on subsystem $\mathcal{H}_\tau$, and $\mathcal{H}_w$, respectively. Equation (89) quantifies how far $\tilde{\rho}_{w\tau}$ is from being a tensor product state where the state on subsystem S_w is maximally mixed. Evidently, if $\tilde{\rho}_w$ is maximally mixed then it will be impossible to train any angle in W .

Let us now analyze the following chain of inequalities valid for the Hilbert-Schmidt distance $D_{HS}(\tilde{\rho}_{w\tau_t}, \tilde{\rho}_{\tau_t} \otimes \frac{\mathbb{1}}{2^m})$ and any choice of S_τ :

$$D_{HS}\left(\tilde{\rho}_{w\tau_t}, \tilde{\rho}_{\tau_t} \otimes \frac{\mathbb{1}}{2^m}\right) \geq \frac{4D_T\left(\tilde{\rho}_{w\tau_t}, \tilde{\rho}_{\tau_t} \otimes \frac{\mathbb{1}}{2^m}\right)^2}{2^m d_\tau} \quad (91)$$

$$\begin{aligned} &\geq \frac{4D_T\left(\tilde{\rho}_w, \frac{\mathbb{1}}{2^m}\right)^2}{2^{m(L-l+2)/2}} \\ &\geq \frac{D_{HS}\left(\tilde{\rho}_w, \frac{\mathbb{1}}{2^m}\right)}{2^{m(L-l+2)/2}}, \end{aligned} \quad (92)$$

with $D_T(A, B) = \text{Tr}[\sqrt{(A - B)^2}]$ the Trace Distance between the Hermitian operators A and B . In the first line we employ the matrix norm equivalence, and we denote as d_τ the dimension of $\mathcal{H}_\tau$. The second line is derived from the fact that the Trace Distance is non-increasing over partial trace [6], and from the fact that $d_\tau \leq 2^{m(L-l)/2} \forall \tau$. Finally, the inequality in (92) employs again the matrix norm equivalence. From Eqs. (92), and (83), we find that the following lower bound holds

$$\text{Var}[\partial_\nu C] \geq \frac{\text{Tr}[\sigma_\nu^2]}{2(2^{2m} - 1)^2(2^m + 1)^{L-l}} \sum_{i \in i_\mathcal{L}} c_i^2 \epsilon_i \left\langle D_{HS}\left(\tilde{\rho}_w, \frac{\mathbb{1}}{2^m}\right) \right\rangle_{V_L}. \quad (93)$$

Following a similar procedure as the one previously employed to compute expectation values over V_R , we can once again leverage the tensor network representation of quantum circuits to algorithmically integrate over each block and compute $\langle D_{HS}(\tilde{\rho}_w, \frac{\mathbb{1}}{2^m}) \rangle_{V_L}$. After considering that all the blocks in V_L which are not in the backpropagated light-cone of W will simplify to identity, we find

$$\left\langle D_{HS}\left(\tilde{\rho}_w, \frac{\mathbb{1}}{2^m}\right) \right\rangle_{V_L} = \sum_{\substack{(k, k') \in k_{\mathcal{L}_B} \\ k' \geq k}} t_{k, k'} \epsilon(\rho_{k, k'}), \quad (94)$$

where $t_{k, k'} \geq 0$, and where $k_{\mathcal{L}_B}$ is the set of k indices whose associated subsystems S_k are in the backward light-cone $\mathcal{L}_B$ of W . Here we defined the function $\epsilon(M) = D_{HS}(M, \text{Tr}(M)\mathbb{1}/d_M)$ where D_{HS} is the Hilbert-Schmidt distance and d_M is the dimension of the matrix M . In addition, $\rho_{k, k'}$ is partial trace of the input state ρ down to the subsystems $S_k S_{k+1} \dots S_{k'}$. In addition, we find that the following inequality holds $\forall k, k'$

$$t_{k, k'} \geq \frac{2^{ml}}{(2^m + 1)^{2l}}. \quad (95)$$

Hence, we have $\text{Var}[\partial_\nu C] \geq G_n(L, l)$, where

$$G_n(L, l) = \frac{2^{m(l+1)-1}}{(2^{2m} - 1)^2(2^m + 1)^{L+l}} \sum_{i \in i_\mathcal{L}} \sum_{\substack{(k, k') \in k_{\mathcal{L}_B} \\ k' \geq k}} c_i^2 \epsilon(\rho_{k, k'}) \epsilon(\hat{O}_i), \quad (96)$$

where we used the fact that $\text{Tr}[\sigma_\nu^2] = 2^m$.

This result can be trivially generalized to the case when $\hat{O}_i$ and $\hat{O}_j$ are of the general form in Eq. (68) such that they can have overlapping support on at most $m/2$ qubits. Then, from (77) it is straightforward to see that $\Delta O_\tau^{ij} = 0$ for all τ, i , and j , as one would always have to compute traces of the form $\text{Tr}[\hat{O}_i^{\mu_i}]$, which vanish since $\hat{O}_i^{\mu_i}$ can be written as a tensor product of Pauli operators. $\square$

Supplementary Note 7: Proof of Theorem 1

The following theorem provides an upper bound on the variance of the partial derivative of a global cost function which can be expressed as the expectation value of an operator of the form

$$O = c_0 \mathbb{1} + \sum_{i=1}^N c_i \hat{O}_{i1} \otimes \hat{O}_{i2} \otimes \cdots \otimes \hat{O}_{i\xi}. \quad (97)$$

Specifically, we consider two cases of interest: (i) When $N = 1$ and each $\hat{O}_{1k}$ is a non-trivial projector ($\hat{O}_{1k}^2 = \hat{O}_{1k} \neq \mathbb{1}$) of rank r_k acting on subsystem S_k , or (ii) When N is arbitrary and $\hat{O}_{ik}$ is traceless with $\text{Tr}[\hat{O}_{ik}^2] \leq 2^m$ (for example, when $\hat{O}_{ik} = \bigotimes_{j=1}^m \sigma_j^\mu$ is a tensor product of Pauli operators $\sigma_j^\mu \in \{\mathbb{1}_j, \sigma_j^x, \sigma_j^y, \sigma_j^z\}$, with at least one $\sigma_j^\mu \neq \mathbb{1}$). In Section 7A we first consider case (i), while in Section 7B we prove Theorem 1 for case (ii).

We now reiterate Theorem 1 for convenience:

Theorem 1. Consider a trainable parameter θ^ν in a block W of the ansatz in Fig. 3 of the main text. Let $\text{Var}[\partial_\nu C]$ be the variance of the partial derivative of a global cost function C (with O given by (97)) with respect to θ^ν . If W_A , W_B of (39), and each block in $V(\theta)$ form a local 2-design, then $\text{Var}[\partial_\nu C]$ is upper bounded by

$$\text{Var}[\partial_\nu C] \leq F_n(L, l). \quad (98)$$

(i) For $N = 1$ and when each $\hat{O}_{1k}$ is a non-trivial projector, then defining $R = \prod_{k=1}^\xi r_k^2$, we have

$$F_n(L, l) = \frac{2^{2m+(2m-1)(L-l)}}{(2^{2m}-1) \cdot 3^{\frac{2n}{m}} \cdot 2^{(2-\frac{3}{m})n}} c_1^2 R. \quad (99)$$

(ii) For arbitrary N and when each $\hat{O}_{ik}$ satisfies $\text{Tr}[\hat{O}_{ik}] = 0$ and $\text{Tr}[\hat{O}_{ik}^2] \leq 2^m$, then

$$F_n(L, l) = \frac{2^{2m(L-l+1)+1}}{3^{\frac{2n}{m}} \cdot 2^{(3-\frac{4}{m})n}} \sum_{i,j=1}^N c_i c_j. \quad (100)$$

A. $N = 1$, and $\hat{O}_{1k}$ is a non-trivial projector

For simplicity, let us now use $\hat{O}_k$ when referring to the operators $\hat{O}_{ik}$. Let us first consider the case when each $\hat{O}_k$ is a non-trivial projector ($\hat{O}_k^2 = \hat{O}_k$) of rank r_k acting on subsystem S_k :

$$r_k := \text{Tr}[\hat{O}_k] = \text{Tr}[\hat{O}_k^2] = \text{rank}[\hat{O}_k]. \quad (101)$$

Proof. In the previous section, we defined V_R as containing all gates in the forward light-cone $\mathcal{L}$ of W and all the gates in the last layer of the ansatz. Hence, we can express $V_R = V_{\mathcal{L}} \otimes V_{\bar{\mathcal{L}}}$, where $V_{\bar{\mathcal{L}}} : \mathcal{H}_{\bar{\mathcal{L}}} \rightarrow \mathcal{H}_{\bar{\mathcal{L}}}$ is given by

$$V_{\bar{\mathcal{L}}} = \bigotimes_{k \in k_{\bar{\mathcal{L}}}} W_{kL}, \quad (102)$$

where $k_{\bar{\mathcal{L}}}$ is the set of k indices whose associated subsystems of qubits S_k are outside of the forward light-cone $\mathcal{L}$ of W . One can always write $|\mathbf{q}\rangle\langle\mathbf{p}| \otimes \mathbb{1}_w$ as a projector onto $H_{\mathcal{L}}$ times a projector onto $H_{\bar{\mathcal{L}}}$:

$$|\mathbf{q}\rangle\langle\mathbf{p}| \otimes \mathbb{1}_w = \left(\bigotimes_{k \in k_{\bar{\mathcal{L}}}} |\mathbf{q}\rangle\langle\mathbf{p}|_k \right) \otimes |\mathbf{q}\rangle\langle\mathbf{p}|_{\mathcal{L}} \otimes \mathbb{1}_w. \quad (103)$$

Combining (49) with Eqs. (102) and (103) leads to

$$\Omega_{qp} = \left(\prod_{k \in k_{\bar{\mathcal{L}}}} \text{Tr}_k \left[|\mathbf{p}\rangle\langle\mathbf{q}|_k W_{kL}^\dagger \hat{O}_k W_{kL} \right] \right) \Omega_{qp}^{\mathcal{L}}, \quad (104)$$

where

$$\Omega_{\mathbf{q}\mathbf{p}}^{\mathcal{L}} = \text{Tr}_{\mathcal{L} \cap \overline{w}} [(|\mathbf{p}\rangle\langle\mathbf{q}| \otimes \mathbb{1}_w) V_{\mathcal{L}} O^{\mathcal{L}} V_{\mathcal{L}}] , \quad (105)$$

and where $O^{\mathcal{L}} = \bigotimes_{k \in k_{\overline{\mathcal{L}}}} \widehat{O}_k$. Here Tr_k indicates the trace over $\mathcal{H}_k$, while $\text{Tr}_{\mathcal{L} \cap \overline{w}}$ is the trace over the Hilbert space associated with the qubits in $S_{\mathcal{L}} \cap S_{\overline{w}}$.

In order to explicitly evaluate the expectation value $\langle \dots \rangle_{V_{\mathbf{R}}}$ in

$$\text{Var}[\partial_{\nu} C] = \frac{2^{m-1} \text{Tr}[\sigma_{\nu}^2]}{(2^{2m} - 1)^2} c_1^2 \sum_{\substack{\mathbf{p}\mathbf{q} \\ \mathbf{p}'\mathbf{q}'}} \left\langle \left(\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}} \Omega_{\mathbf{q}'\mathbf{p}'}] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}]}{2^m} \right) \right\rangle_{V_{\mathbf{R}}} \left\langle \Delta \Psi_{\mathbf{p}\mathbf{q}}^{\mathbf{p}'\mathbf{q}'} \right\rangle_{V_{\mathbf{L}}} , \quad (106)$$

we use the fact that the blocks in $V(\boldsymbol{\theta})$ are independent, and hence $\langle \dots \rangle_{V_{\mathbf{R}}} = \langle \dots \rangle_{V_{\overline{\mathcal{L}}}, V_{\mathcal{L}}}$. Then, we have

$$\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}} \Omega_{\mathbf{q}'\mathbf{p}'}] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}]}{2^m} \right\rangle_{V_{\mathbf{R}}} = \left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^{\mathcal{L}} \Omega_{\mathbf{q}'\mathbf{p}'}^{\mathcal{L}}] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^{\mathcal{L}}] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^{\mathcal{L}}]}{2^m} \right\rangle_{V_{\mathcal{L}}} \left(\prod_{k/S_k \subset S_{\overline{\mathcal{L}}}} \langle \Omega_k \rangle_{W_{kL}} \right) , \quad (107)$$

with

$$\Omega_k = \text{Tr}_k [|\mathbf{p}\rangle\langle\mathbf{q}|_k W_{kL}^{\dagger} \widehat{O}_k W_{kL}] \text{Tr}_k [|\mathbf{p}'\rangle\langle\mathbf{q}'|_k W_{kL}^{\dagger} \widehat{O}_k W_{kL}] . \quad (108)$$

Let us first compute the expectation value of each Ω_k . From Lemma 3, we have:

$$\begin{aligned} \langle \Omega_k \rangle_{W_{kL}} &= \frac{r_k}{2^{2m} - 1} \left(\left(r_k - \frac{1}{2^m} \right) \delta_{(\mathbf{p}, \mathbf{q})_{S_k}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_k}} + \left(1 - \frac{r_k}{2^m} \right) \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \right) \\ &\leq \frac{1}{2^{2m} - 1} r_k^2 \left(\delta_{(\mathbf{p}, \mathbf{q})_{S_k}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_k}} + \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \right) , \end{aligned} \quad (109)$$

where in the inequality we have dropped the negative terms and used the fact that $r_k \leq r_k^2$. Then, we have

$$\prod_{k \in k_{\overline{\mathcal{L}}}} \langle \Omega_k \rangle_{W_{kL}} \leq \frac{1}{(2^{2m} - 1)^{\xi_{\overline{\mathcal{L}}}}} \prod_{k \in k_{\overline{\mathcal{L}}}} r_k^2 \left(\delta_{(\mathbf{p}, \mathbf{q})_{S_k}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_k}} + \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \right) .$$

Combining this result with Eq. (106) leads to the upper bound

$$\begin{aligned} \text{Var}[\partial_{\nu} C] &\leq \frac{2^{m-1} \text{Tr}[\sigma_{\nu}^2]}{(2^{2m} - 1)^2 (2^{2m} - 1)^{\xi_{\overline{\mathcal{L}}}}} c_1^2 \sum_{\substack{\mathbf{p}\mathbf{q} \\ \mathbf{p}'\mathbf{q}'}} \left(\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^{\mathcal{L}} \Omega_{\mathbf{q}'\mathbf{p}'}^{\mathcal{L}}] - \frac{\text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^{\mathcal{L}}] \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^{\mathcal{L}}]}{2^m} \right\rangle_{V_{\mathcal{L}}} \right. \\ &\quad \times \left. \left(\prod_{k \in k_{\overline{\mathcal{L}}}} r_k^2 \left(\delta_{(\mathbf{p}, \mathbf{q})_{S_k}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_k}} + \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \right) \right) \left\langle \Delta \Psi_{\mathbf{p}\mathbf{q}}^{\mathbf{p}'\mathbf{q}'} \right\rangle_{V_{\mathbf{L}}} \right) . \end{aligned} \quad (110)$$

As discussed in the previous section, one can compute the expectation values in Eq. (110) by systematically integrating over each block over the unitary group with the respect to the Haar measure. From Eq. (76) we can find

$$\begin{aligned} \text{Var}[\partial_{\nu} C] &\leq \frac{2^{m-1} \text{Tr}[\sigma_{\nu}^2]}{(2^{2m} - 1)^2 (2^{2m} - 1)^{\xi_{\overline{\mathcal{L}}}}} c_1^2 \sum_{\substack{\mathbf{p}\mathbf{q} \\ \mathbf{p}'\mathbf{q}'}} \sum_{\tau} \left(t_{\tau} \delta_{(\mathbf{p}, \mathbf{q})_{S_{\overline{\tau}}}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_{\overline{\tau}}}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_{\tau}}} \Delta O_{\tau} \right. \\ &\quad \times \left. \left(\prod_{k \in k_{\overline{\mathcal{L}}}} r_k^2 \left(\delta_{(\mathbf{p}, \mathbf{q})_{S_k}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_k}} + \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \right) \right) \left\langle \Delta \Psi_{\mathbf{p}\mathbf{q}}^{\mathbf{p}'\mathbf{q}'} \right\rangle_{V_{\mathbf{L}}} \right) . \end{aligned} \quad (111)$$

Note that by expanding the product $\prod_{k \in k_{\overline{\mathcal{L}}}} \left(\delta_{(\mathbf{p}, \mathbf{q})_{S_k}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_k}} + \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \right)$ $\delta_{(\mathbf{p}, \mathbf{q})_{S_{\overline{\tau}}}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_{\overline{\tau}}}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_{\tau}}}$, one obtains a sum of $2^{\xi_{\overline{\mathcal{L}}}}$ terms, and according to Eqs. (86)–(89), each term in the summation leads to a Hilbert-Schmidt distances between two quantum states. Then, since $D_{HS}(\rho_1, \rho_2) \leq 2$ for any pair of states ρ_1, ρ_2 , we find

$$\sum_{\substack{\mathbf{p}\mathbf{q} \\ \mathbf{p}'\mathbf{q}'}} \delta_{(\mathbf{p}, \mathbf{q})_{S_{\overline{\tau}}}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_{\overline{\tau}}}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_{\tau}}} \left(\prod_{k \in k_{\overline{\mathcal{L}}}} \left(\delta_{(\mathbf{p}, \mathbf{q})_{S_k}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_k}} + \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \right) \right) \left\langle \Delta \Psi_{\mathbf{p}\mathbf{q}}^{\mathbf{p}'\mathbf{q}'} \right\rangle_{V_{\mathbf{L}}} \leq 2 \cdot 2^{\xi_{\overline{\mathcal{L}}}} . \quad (112)$$

Replacing this result in (111) leads to

$$\text{Var}[\partial_\nu C] \leq \frac{2^m 2^{\xi_{\mathcal{L}}} \text{Tr}[\sigma_\nu^2]}{(2^m - 1)^2 (2^{2m} - 1)^{\xi_{\mathcal{L}}}} c_1^2 \left(\prod_{k \in k_{\mathcal{L}}} r_k^2 \right) \sum_{\tau} t_{\tau} \Delta O_{\tau}. \quad (113)$$

Next, we consider the terms ΔO_{τ} . From Eq. (77) we can show that

$$\Delta O_{\tau} \leq \prod_{k \in k_{\mathcal{L}}} r_k^2, \quad (114)$$

so that (113) becomes

$$\text{Var}[\partial_\nu C] \leq \frac{2^m 2^{\xi_{\mathcal{L}}} \text{Tr}[\sigma_\nu^2]}{(2^{2m} - 1)^2 (2^{2m} - 1)^{\xi_{\mathcal{L}}}} c_1^2 R \sum_{\tau} t_{\tau}, \quad (115)$$

where $R = \prod_{k=1}^{\xi} r_k^2$.

Let us finally show that $\sum_{\tau} t_{\tau} \leq 2 \forall l, L$. We recall from Eq. (106) that the coefficients t_{τ} are obtained by integrating each block in $V(\theta)$ over the unitary group with the respect to the Haar measure. Each time a block is integrated one obtains four new tensors weighted by the coefficients η_i , with $i = 1, \dots, 4$ such that $\sum_{\mu=1}^4 |\eta_{\mu}| \leq 1$ for all m . Consider now the average $\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] \right\rangle_{V_{\mathbf{R}}}$, once all the blocks have been integrated we find

$$\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \Omega_{\mathbf{q}'\mathbf{p}'}^j] \right\rangle_{V_{\mathbf{R}}} = \sum_{\mu_1=1}^4 \eta_{\mu_1} \left(\sum_{\mu_2=1}^4 \eta_{\mu_2} \left(\dots \left(\sum_{\mu_m=1}^4 \eta_{\mu_m} \right) \right) \right) T_{\mu_1, \mu_2, \dots, \mu_m}(O), \quad (116)$$

where μ_m is the number of averaged blocks, and where $T_{\mu_1, \mu_2, \dots, \mu_m}(O)$ is a tensor contraction of O . A similar equation can be obtained for $\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]] \right\rangle_{V_{\mathbf{R}}}$ as

$$\left\langle \text{Tr}[\Omega_{\mathbf{q}\mathbf{p}}^i \text{Tr}[\Omega_{\mathbf{q}'\mathbf{p}'}^j]] \right\rangle_{V_{\mathbf{R}}} = \sum_{\mu'_1=1}^4 \eta_{\mu'_1} \left(\sum_{\mu'_2=1}^4 \eta_{\mu'_2} \left(\dots \left(\sum_{\mu'_m=1}^4 \eta_{\mu'_m} \right) \right) \right) T_{\mu'_1, \mu'_2, \dots, \mu'_m}(O), \quad (117)$$

so that

$$\sum_{\tau} t_{\tau} = \sum_{\mu_1=1}^4 \eta_{\mu_1} \left(\sum_{\mu_2=1}^4 \eta_{\mu_2} \left(\dots \left(\sum_{\mu_m=1}^4 \eta_{\mu_m} \right) \right) \right) - \frac{1}{2^m} \sum_{\mu'_1=1}^4 \eta_{\mu'_1} \left(\sum_{\mu'_2=1}^4 \eta_{\mu'_2} \left(\dots \left(\sum_{\mu'_m=1}^4 \eta_{\mu'_m} \right) \right) \right). \quad (118)$$

Taking the absolute value on both side and using the fact that $\sum_{\mu=1}^4 |\eta_{\mu}| \leq 1$, one gets $\sum_{\tau} t_{\tau} = |\sum_{\tau} t_{\tau}| \leq 1 + \frac{1}{2^m} \leq 2$.

Therefore, by using $\text{Tr}[\sigma_\nu^2] = 2^m$, we have

$$\text{Var}[\partial_\nu C] \leq \frac{2^{2m + \frac{n}{m} + l - L}}{(2^{2m} - 1)^2 (2^{2m} - 1)^{\frac{n}{m} + l - L - 1}} c_1^2 R.$$

Moreover, we also find that

$$\frac{2^{2m + \frac{n}{m} + l - L}}{(2^{2m} - 1)^2 (2^{2m} - 1)^{\frac{n}{m} + l - L - 1}} c_1^2 R = \frac{2^{2m}}{2^{2m} - 1} \cdot \left(2^{2m-1} - \frac{1}{2} \right)^{L-l} \cdot \frac{2^{\frac{n}{m}}}{2^{2n} (1 - 2^{-2m})^{\frac{n}{m}}} c_1^2 R, \quad (119)$$

and since $2^{2m-1} - \frac{1}{2} < 2^{2m-1}$ and $1 \leq \frac{1}{1 - 2^{-2m}} \leq \frac{4}{3}$, $\text{Var}[\partial_\nu C]$ can be upper bounded as $\text{Var}[\partial_\nu C] < F_n(L, l)$, where

$$F_n(L, l) = \frac{2^{2m + (2m-1)(L-l)}}{(2^{2m} - 1) \cdot 3^{\frac{n}{m}} \cdot 2^{(2 - \frac{3}{m})n}} c_1^2 R. \quad (120)$$

□

B. Arbitrary N and $\widehat{O}_{ik}$ is a traceless operators such that $\text{Tr}[\widehat{O}_{ik}^2] \leq 2^m$

We now consider the case when $\widehat{O}_{ik}$ is traceless. Specifically, we assume

$$\text{Tr}[\widehat{O}_{ik}] = 0, \quad \text{and} \quad \text{Tr}[\widehat{O}_{ik}^2] \leq 2^m. \quad (121)$$

Note that if $\widehat{O}_{ik}$ is a tensor product of Pauli operators then $\text{Tr}[\widehat{O}_{ik}^2] = 2^m$.

Proof. Let us first analyze the case of $N = 1$. Combining Lemma 3 with Eqs. (107), and (108), we find

$$\langle \Omega_k \rangle_{W_{kL}} \leq \frac{2^m}{2^{2m}-1} \left(\delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} - \frac{1}{2^m} \delta_{(\mathbf{p}, \mathbf{q})_{S_k}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_k}} \right) \leq \frac{2^m}{2^{2m}-1} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}}, \quad (122)$$

which yields

$$\prod_{k \in k_{\overline{\mathcal{L}}}} \langle \Omega_k \rangle_{W_{kL}} \leq \frac{2^{\xi_{\overline{\mathcal{L}}}m}}{(2^{2m}-1)^{\xi_{\overline{\mathcal{L}}}}} \prod_{k \in k_{\overline{\mathcal{L}}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}}. \quad (123)$$

Therefore, from (111), and (123), we can obtain

$$\text{Var}[\partial_\nu C] \leq \frac{2^{2m-1}}{(2^{2m}-1)^2(2^{2m}-1)^{\xi_{\overline{\mathcal{L}}}}} c_1^2 \cdot \frac{2^{\xi_{\overline{\mathcal{L}}}m}}{(2^{2m}-1)^{\xi_{\overline{\mathcal{L}}}}} \quad (124)$$

$$\times \sum_{\substack{\mathbf{p}\mathbf{q} \\ \mathbf{p}'\mathbf{q}'}} \sum_{\tau} t_{\tau} \Delta O_{\tau} \left(\prod_{k \in k_{\overline{\mathcal{L}}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \right) \delta_{(\mathbf{p}, \mathbf{q})_{S_{\overline{\mathcal{T}}}}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_{\overline{\mathcal{T}}}}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_{\tau}}} \left\langle \Delta \Psi_{\mathbf{p}\mathbf{q}}^{\mathbf{p}'\mathbf{q}'} \right\rangle_{V_L}. \quad (125)$$

Let us now consider the terms ΔO_{τ} from (77). It is straightforward to see that $\Delta O_{\tau} = 0$ if $S_{\tau_z} \neq \emptyset$, due to the fact that $\text{Tr}[\widehat{O}_{ik}] = 0$. Hence, let us define τ_p as the set of indexes such that $S_{\tau_z} = \emptyset$. Then, we have

$$\Delta O_{\tau} \leq \begin{cases} 2^{m(L-l+1)} & (\tau \in \tau_p) \\ 0 & (\tau \notin \tau_p) \end{cases}. \quad (126)$$

The latter implies

$$\sum_{\tau} t_{\tau} \Delta O_{\tau} \leq \sum_{\tau \in \tau_p} t_{\tau} \Delta O_{\tau} \leq 2 \times 2^{m(L-l+1)}, \quad (127)$$

where we used the fact that $\sum t_{\tau} \leq 2$ (see Eq. (118) in the previous section).

According to Eqs. (86)–(89), the product $\left(\prod_{k \in k_{\overline{\mathcal{L}}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_k}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_k}} \right) \delta_{(\mathbf{p}, \mathbf{q})_{S_{\overline{\mathcal{T}}}}} \delta_{(\mathbf{p}', \mathbf{q}')_{S_{\overline{\mathcal{T}}}}} \delta_{(\mathbf{p}, \mathbf{q}')_{S_{\tau}}} \delta_{(\mathbf{p}', \mathbf{q})_{S_{\tau}}}$, leads to a Hilbert-Schmidt distances between two quantum states, which is always upper bounded by 2. We then find

$$\text{Var}[\partial_\nu C] \leq \frac{2^{2m+1}}{(2^{2m}-1)^2(2^{2m}-1)^{\xi_{\overline{\mathcal{L}}}}} c_1^2 \cdot \frac{2^{\xi_{\overline{\mathcal{L}}}m}}{(2^{2m}-1)^{\xi_{\overline{\mathcal{L}}}}} \cdot 2^{m(L-l+1)}. \quad (128)$$

Since $n = m(L-l+1) + m\xi_{\overline{\mathcal{L}}}$, we can rewrite (128) as

$$\text{Var}[\partial_\nu C] \leq \frac{2^{2m+n+1}(2^{2m}-1)^{2(L-l)}}{2^{4n}(1-2^{-2m})^{\frac{2n}{m}}} c_1^2. \quad (129)$$

Finally, noting that $2^{2m}-1 \leq 2^{2m}$ and $1 \leq \frac{1}{1-2^{-2m}} \leq \frac{4}{3}$, we have

$$\text{Var}[\partial_\nu C] \leq \frac{2^{2m(L-l+1)+1}}{3^{\frac{2n}{m}} 2^{(3-\frac{4}{m})n}} c_1^2. \quad (130)$$

Let us now consider the arbitrary N case, where the operator O can be expressed as

$$O = c_0 \mathbb{1} + \sum_{i=1}^N c_i O_i, \quad (131)$$

and where each O_i can be expressed as $O_i = \widehat{O}_{i1} \otimes \widehat{O}_{i2} \otimes \cdots \otimes \widehat{O}_{i\xi}$ with $\widehat{O}_{ik}$ satisfying (121).

Here it is convinient to define $C_i = \text{Tr}[O_i V(\boldsymbol{\theta}) \rho V^\dagger(\boldsymbol{\theta})]$, so that $C = c_0 + \sum_i c_i C_i$. When computing the variance we now have to consider the cross terms arising from C_i , and C_j with $i \neq j$:

$$\text{Var}[\partial_\nu C] = \langle (\partial_\nu C)^2 \rangle_V = \sum_i c_i^2 \langle (\partial_\nu C_i)^2 \rangle_V + \sum_{i \neq j} c_i c_j \langle \partial_\nu C_i \partial_\nu C_j \rangle_V. \quad (132)$$

First, let us remark that

$$\text{Tr}[\widehat{O}_{ik}] = 0, \quad \text{and} \quad \text{Tr}[\widehat{O}_{ik} \widehat{O}_{jk'}] \leq 2^m, \quad \forall i, j, k, k' \quad \text{and} \quad \begin{cases} \Delta O_\tau^{ij} \leq 2^{m(L-l+1)} & (\tau \in \tau_p) \\ \Delta O_\tau^{ij} = 0 & (\tau \notin \tau_p) \end{cases}, \quad (133)$$

where ΔO_τ^{ij} was defined in (77), and where $\text{Tr}[\widehat{O}_{ik} \widehat{O}_{jk'}] \leq 2^m$ follows from the fact that $\text{Tr}[\widehat{O}_{ik}^2] \leq 2^m$ and $\text{Tr}[\widehat{O}_{jk'}^2] \leq 2^m$. Hence, it is straightforward to see that the upper bound in Eq. (130) also holds for the cross terms in (132), and we then have

$$\text{Var}[\partial_\nu C] \leq F_n(L, l). \quad (134)$$

where

$$F_n(L, l) = \frac{2^{2m(L-l+1)+1}}{3^{\frac{2n}{m}} 2^{(3-\frac{4}{m})n}} \sum_{ij} c_i c_j. \quad (135)$$

□

Supplementary Note 8: Proofs of Corollaries

A. Proof of Corollary 1

Corollary 1. Consider the function $F_n(L, l)$.

- (i) Let $N = 1$ and let each $\widehat{O}_{1k}$ be a non-trivial projector, as in case (i) of Theorem 1. If $c_1^2 R \in \mathcal{O}(2^n)$ and if the number of layers $L \in \mathcal{O}(\text{poly}(\log(n)))$, then

$$F_n(L, l) \in \mathcal{O}\left(2^{-(1-\frac{1}{m}\log_2 3)n}\right), \quad (136)$$

which implies that $\text{Var}[\partial_\nu C]$ is exponentially vanishing in n if $m \geq 2$.

- (ii) Let N be arbitrary, and let each $\widehat{O}_{ik}$ satisfy $\text{Tr}[\widehat{O}_{ik}] = 0$ and $\text{Tr}[\widehat{O}_{ik}^2] \leq 2^m$, as in case (ii) of Theorem 1. If $N \in \mathcal{O}(2^n)$, $c_i \in \mathcal{O}(1)$, and if the number of layers $L \in \mathcal{O}(\text{poly}(\log(n)))$, then

$$F_n(L, l) \in \mathcal{O}\left(\frac{1}{2^{(1-\frac{1}{m})n}}\right), \quad (137)$$

which implies that $\text{Var}[\partial_\nu C]$ is exponentially vanishing in n if $m \geq 2$.

Let us first consider case (i) in Theorem 1.

Proof. Let us assume $L \in \mathcal{O}(\text{poly}(\log(n)))$, so that we have $2^{(2m-1)(L-l)} \in \mathcal{O}(2^{\text{poly}(\log(n))})$. Here, note that

$$\lim_{n \rightarrow \infty} \left| \frac{\text{poly}(\log(n))}{n} \right| = 0,$$

which means that n grows faster than the any polylogarithmic functions of n . Therefore, we can write $\mathcal{O}(\text{poly}(\log(n))) \subset \mathcal{O}(n)$. For a nonzero constant a , we can write $\mathcal{O}(|a|n) = \mathcal{O}(n)$, and we can choose $a = \frac{1}{m} \log_2(9/8)$. Therefore, we can also write $\mathcal{O}(2^{\text{poly}(\log(n))}) \subset \mathcal{O}\left((9/8)^{\frac{n}{m}}\right)$. In addition, if we have $c_1^2 R \in \mathcal{O}(2^n)$, from (120), one can obtain

$$F_n(L, l) \in \frac{1}{3^{\frac{n}{m}} 2^{(2-\frac{3}{m})n}} \mathcal{O}\left((9/8)^{\frac{n}{m}}\right) \mathcal{O}(2^n) = \mathcal{O}\left(\frac{1}{2^{(1-\frac{1}{m}\log_2 3)n}}\right), \quad (138)$$

where $1 - \frac{1}{m} \log_2 3 > 0$ for $m \geq 2$. Hence the upper bound of $\text{Var}[\partial_\nu C_G]$ exponentially vanishing when $m \geq 2$. □

Let us now consider case (ii) in Theorem 1.

Proof. We now have $2^{2m(L-l+1)+1} \in \mathcal{O}(2^{\text{poly}(\log(n))}) \subset \mathcal{O}\left((9/8)^{\frac{n}{m}}\right)$. We have from Eq. (135) that if $c_i, c_j \in \mathcal{O}(1)$, the following bound holds

$$F_n(L, l) \in \frac{1}{3^{\frac{2n}{m}} 2^{(3-\frac{4}{m})n}} \mathcal{O}\left((9/8)^{\frac{n}{m}}\right) \mathcal{O}(2^{2n}) = \mathcal{O}\left(\frac{1}{2^{(1-\frac{1}{m})n}}\right).$$

Hence, the upper bound of $\text{Var}[\partial_\nu C_G]$ exponentially vanishing when $m \geq 2$. $\square$

B. Proof of Corollary 2

Corollary 2. Consider the function $F_n(L, l)$. Let O be an operator of the form (68), as in Theorem 2. If at least one term $c_i^2 \epsilon(\rho_{k,k'}) \epsilon(\widehat{O}_i)$ in the sum in (70) vanishes no faster than $\Omega(1/\text{poly}(n))$, and if the number of layers L is $\mathcal{O}(\log(n))$, then

$$G_n(L, l) \in \Omega\left(\frac{1}{\text{poly}(n)}\right). \quad (139)$$

On the other hand, if at least one term $c_i^2 \epsilon(\rho_{k,k'}) \epsilon(\widehat{O}_i)$ in the sum in (70) vanishes no faster than $\Omega(1/2^{\text{poly}(\log(n))})$, and if the number of layers is $\mathcal{O}(\text{poly}(\log(n)))$, then

$$G_n(L, l) \in \Omega\left(\frac{1}{2^{\text{poly}(\log(n))}}\right). \quad (140)$$

Proof. Let us assume that at least one term $c_i^2 \epsilon(\rho_{k,k'}) \epsilon(\widehat{O}_i)$ in (70) vanishes no faster than $\Omega(1/\text{poly}(n))$. Then, if we also assume $L \in \mathcal{O}(\log(n))$, we have $(2^m + 1)^{-(L+l)} \in \Omega(1/\text{poly}(n))$. The latter implies

$$G_n(L, l) \in \Omega\left(\frac{1}{\text{poly}(n)}\right) \Omega\left(\frac{1}{\text{poly}(n)}\right) = \Omega\left(\frac{1}{\text{poly}(n)}\right).$$

On the other hand, if at least one term $c_i^2 \epsilon(\rho_{k,k'}) \epsilon(\widehat{O}_i)$ in (70) vanishes no faster than $\Omega(1/2^{\text{poly}(\log(n))})$, and if $L \in \mathcal{O}(\text{poly}(\log(n)))$, we have $(2^m + 1)^{-(L+l)} \in \Omega(1/2^{\text{poly}(\log(n))})$. Therefore, we obtain

$$G_n(L, l) \in \Omega\left(\frac{1}{2^{\text{poly}(\log(n))}}\right) \Omega\left(\frac{1}{2^{\text{poly}(\log(n))}}\right) = \Omega\left(\frac{1}{2^{\text{poly}(\log(n))}}\right).$$

$\square$

Supplementary Note 9: Faithfulness of local cost function for Quantum autoencoder

Recall that the global cost function (C'_G) and local cost function (C'_L) for the quantum autoencoder are defined as

$$\begin{aligned} C'_G &= 1 - \text{Tr} [|0\rangle\langle 0| \rho_B^{\text{out}}] \\ C'_L &= 1 - \frac{1}{n_B} \sum_{j=1}^{n_B} \text{Tr} \left[\left(|0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}} \right) \rho_B^{\text{out}} \right]. \end{aligned}$$

We relate the C'_L and C'_G cost functions in the following proposition. Because this proposition establishes that C'_L and C'_G vanish under the same conditions, and because C'_G is faithful [7], this in turn proves that C'_L is a faithful cost function.

Proposition 3. The cost functions for the quantum autoencoder satisfy

$$C'_L \leq C'_G \leq n_B C'_L. \quad (141)$$

Proof. Let us first prove $C'_L \leq C'_G$. Given the state ρ_B^{out} , we can define E_j as the event of qubit j being measured on the $|0\rangle_j$ state, such that the probability of E_j is given by $\Pr(E_j) = \text{Tr}[\hat{O}_j^L \rho_B^{\text{out}}]$, where $\hat{O}_j^L = |0\rangle\langle 0|_j \otimes \mathbb{1}_{\bar{j}}$. Then, we can write

$$C'_G = 1 - \Pr\left(\bigcap_{j=1}^{n_B} E_j\right) \quad (142)$$

Similarly, for the local cost function, we have

$$C'_L = 1 - \frac{1}{n_B} \sum_{j=1}^{n_B} \Pr(E_j). \quad (143)$$

Then, it is known that for any set of events $\{E_1, \dots, E_{n_B}\}$, the following property always holds

$$1 - \Pr\left(\bigcap_{j=1}^{n_B} E_j\right) \geq \frac{1}{n_B} \sum_{j=1}^{n_B} (1 - \Pr(E_j)). \quad (144)$$

From the definition of C'_G and C'_L in Eqs. (142)–(143), and from Eq. (144) we have

$$C'_L \leq C'_G. \quad (145)$$

Next, let us prove $C'_G \leq n_B C'_L$. Since for any set of events $\{E_1, \dots, E_{n_B}\}$, we have

$$1 - \Pr\left(\bigcap_{j=1}^{n_B} E_j\right) \leq \sum_{j=1}^{n_B} (1 - \Pr(E_j)). \quad (146)$$

By definition, we finally have

$$C'_G \leq n_B C'_L. \quad (147)$$

Combining Eqs. (145) and (147), we get (141), which indicates $C'_L = 0 \iff C'_G = 0$. $\square$

-
- [1] Benoît Collins and Piotr Śniady, “Integration with respect to the haar measure on unitary, orthogonal and symplectic group,” *Communications in Mathematical Physics* **264**, 773–795 (2006).
 - [2] Zbigniew Puchała and Jarosław Adam Miszczyk, “Symbolic integration with respect to the haar measure on the unitary groups,” *Bulletin of the Polish Academy of Sciences Technical Sciences* **65**, 21–27 (2017).
 - [3] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine, “Exact and approximate unitary 2-designs and their application to fidelity estimation,” *Physical Review A* **80**, 012304 (2009).
 - [4] R. E. A. C. Paley and A. Zygmund, “A note on analytic functions in the unit circle,” *Math. Proc. Camb. Phil. Soc.* **28**, 266 (1932).
 - [5] Motohisa Fukuda, Robert König, and Ion Nechita, “RTNI—a symbolic integrator for haar-random tensor networks,” *Journal of Physics A: Mathematical and Theoretical* **52**, 425303 (2019).
 - [6] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition*, 10th ed. (Cambridge University Press, New York, NY, USA, 2011).
 - [7] J. Romero, J. P. Olson, and A. Aspuru-Guzik, “Quantum autoencoders for efficient compression of quantum data,” *Quantum Science and Technology* **2**, 045001 (2017).